



INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification ⁶ : C12N 15/82, 9/02, 5/10, G01N 33/50		A2	(11) International Publication Number: WO 99/55888
			(43) International Publication Date: 4 November 1999 (04.11.99)
(21) International Application Number: PCT/US99/08746		(81) Designated States: AE, AL, AU, BA, BB, BG, BR, CA, CN, CU, CZ, EE, GD, GE, HR, HU, ID, IL, IN, IS, JP, KP, KR, LC, LK, LR, LT, LV, MG, MK, MN, MX, NO, NZ, PL, RO, SG, SI, SK, SL, TR, TT, UA, US, UZ, VN, YU, ZA, ARIPO patent (GH, GM, KE, LS, MW, SD, SL, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).	
(22) International Filing Date: 21 April 1999 (21.04.99)			
(30) Priority Data: 60/083,042 24 April 1998 (24.04.98) US			
(71) Applicant (for all designated States except US): E.I. DU PONT DE NEMOURS AND COMPANY [US/US]; 1007 Market Street, Wilmington, DE 19898 (US).			
(72) Inventors; and (75) Inventors/Applicants (for US only): CAHOON, Rebecca, E. [US/US]; 2331 West 18th Street, Wilmington, DE 19806 (US). KINNEY, Anthony, J. [GB/US]; 609 Lore Avenue, Wilmington, DE 19809 (US). VOLLMER, Steven, J. [US/US]; 1019 Tweedbrook Road, Wilmington, DE 19810 (US). WENG, Zude [CN/US]; 909C Cloister Road, Wilmington, DE 19809 (US).			
(74) Agent: MAJARIAN, William, R.; E.I. du Pont de Nemours and Company, Legal Patent Records Center, 1007 Market Street, Wilmington, DE 19898 (US).			
(54) Title: CAROTENOID BIOSYNTHESIS ENZYMES			
(57) Abstract			
<p>This invention relates to an isolated nucleic acid fragment encoding a carotenoid biosynthetic enzyme. The invention also relates to the construction of a chimeric gene encoding all or a portion of the carotenoid biosynthetic enzyme, in sense or antisense orientation, wherein expression of the chimeric gene results in production of altered levels of the carotenoid biosynthetic enzyme in a transformed host cell.</p>			
<pre> ** SEQ ID NO:13 NASVAA--TTTAPALAFRRARPGTC-----LVFFP--RASAAVAASTVTSPTWR- SEQ ID NO:14 NATCAVLCCTATSAALKKRVFPDQAGLFFGGRALSHLV--TFSEVIRADLNSHVDN- SEQ ID NO:06 NAS--LIQCATS-----LSAVITTTATKTRKRLACSLANVSDH- SEQ ID NO:08 MA--AT--SCALVSALVCKR--G-----PSCOR--AAAGVVRCSLQSKVSDH- 1 60 ***** SEQ ID NO:13 --GASQLFFPEFHTGAPKLVATIGAGLAGHSTAVELLDOGREVLTYESAPFTGGKVG SEQ ID NO:14 STNAFGLFFPEFHTGAPKLVATIGAGLAGHSTAVELLDOGREVLTYESAPFTGGKVG SEQ ID NO:06 STNAFGLFFPEFHTGAPKLVATIGAGLAGHSTAVELLDOGREVLTYESAPFTGGKVG SEQ ID NO:08 AINAFGLFFPEFHTGAPKLVATIGAGLAGHSTAVELLDOGREVLTYESAPFTGGKVG 61 120 ***** SEQ ID NO:13 SFVDGKGNHIEGLVFFGCTSNLFLAKGVAGDNLVKNHTTFFVNGGTICGLDFFP SEQ ID NO:14 SFVDGKGNHIEGLVFFGCTSNLFLAKGVAGDNLVKNHTTFFVNGGTICGLDFFP SEQ ID NO:06 SFVDGKGNHIEGLVFFGCTSNLFLAKGVAGDNLVKNHTTFFVNGGTICGLDFFP SEQ ID NO:08 SFVDGKGNHIEGLVFFGCTSNLFLAKGVAGDNLVKNHTTFFVNGGTICGLDFFP 121 180 ***** SEQ ID NO:13 PVGAPLNGIQAFLTQKLVYDKARNAVALALSPPVVALVDPDGLQGVRLDDISFSDW SEQ ID NO:14 PVGAPLNGIQAFLTQKLVYDKARNAVALALSPPVVALVDPDGLQGVRLDDISFSDW SEQ ID NO:06 PVGAPLNGIQAFLTQKLVYDKARNAVALALSPPVVALVDPDGLQGVRLDDISFSDW SEQ ID NO:08 PVGAPLNGIQAFLTQKLVYDKARNAVALALSPPVVALVDPDGLQGVRLDDISFSDW 181 240 ***** SEQ ID NO:13 FNSGGTRESITWNPVYALGFIQDCHISARCLTITFLPATKTEABLLRLKGSFV SEQ ID NO:14 FNSGGTRESITWNPVYALGFIQDCHISARCLTITFLPATKTEABLLRLKGSFV SEQ ID NO:06 FNSGGTRESITWNPVYALGFIQDCHISARCLTITFLPATKTEABLLRLKGSFV SEQ ID NO:08 FNSGGTRESITWNPVYALGFIQDCHISARCLTITFLPATKTEABLLRLKGSFV 241 300 ***** SEQ ID NO:13 YLGGPIKRYITDGGPHLNGCRVLYEYKSPDGETTVGGLITKATSEIISKADATVAA SEQ ID NO:14 YLGGPIKRYITDGGPHLNGCRVLYEYKSPDGETTVGGLITKATSEIISKADATVAA SEQ ID NO:06 YLGGPIKRYITDGGPHLNGCRVLYEYKSPDGETTVGGLITKATSEIISKADATVAA SEQ ID NO:08 YLGGPIKRYITDGGPHLNGCRVLYEYKSPDGETTVGGLITKATSEIISKADATVAA 301 360 ***** SEQ ID NO:13 CDPGKIKALLPSNREWFONITVLDOVPPVVGIVGNGVTELOKESRQLQAVGL SEQ ID NO:14 CDPGKIKALLPSNREWFONITVLDOVPPVVGIVGNGVTELOKESRQLQAVGL SEQ ID NO:06 CDPGKIKALLPSNREWFONITVLDOVPPVVGIVGNGVTELOKESRQLQAVGL SEQ ID NO:08 CDPGKIKALLPSNREWFONITVLDOVPPVVGIVGNGVTELOKESRQLQAVGL 361 420 ***** SEQ ID NO:13 ONLLYTADAFSCFDLALSSPDTTIEGQSLIQAFLTPODPENFLPHEEISKVKQGV SEQ ID NO:14 ONLLYTADAFSCFDLALSSPDTTIEGQSLIQAFLTPODPENFLPHEEISKVKQGV SEQ ID NO:06 ONLLYTADAFSCFDLALSSPDTTIEGQSLIQAFLTPODPENFLPHEEISKVKQGV SEQ ID NO:08 ONLLYTADAFSCFDLALSSPDTTIEGQSLIQAFLTPODPENFLPHEEISKVKQGV 421 480 ***** SEQ ID NO:13 VELFPEERGLVTVSSVKIGQSLYRAGPDGFPDQPTFVKNITLAGSYTKDDYIDSM SEQ ID NO:14 VELFPEERGLVTVSSVKIGQSLYRAGPDGFPDQPTFVKNITLAGSYTKDDYIDSM SEQ ID NO:06 VELFPEERGLVTVSSVKIGQSLYRAGPDGFPDQPTFVKNITLAGSYTKDDYIDSM SEQ ID NO:08 VELFPEERGLVTVSSVKIGQSLYRAGPDGFPDQPTFVKNITLAGSYTKDDYIDSM 481 540 ***** SEQ ID NO:13 ECATLSCRTAAVTCAGGELLALAKGLA--IDGGRALGVVPLDAS SEQ ID NO:14 ECATLSCRTAAVTCAGGELLALAKGLA--IDGGRALGVVPLDAS SEQ ID NO:06 ECATLSCRTAAVTCAGGELLALAKGLA--IDGGRALGVVPLDAS SEQ ID NO:08 ECATLSCRTAAVTCAGGELLALAKGLA--IDGGRALGVVPLDAS 541 599 </pre>			

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece	ML	Mali	TR	Turkey
BG	Bulgaria	HU	Hungary	MN	Mongolia	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MR	Mauritania	UA	Ukraine
BR	Brazil	IL	Israel	MW	Malawi	UG	Uganda
BY	Belarus	IS	Iceland	MX	Mexico	US	United States of America
CA	Canada	IT	Italy	NE	Niger	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NL	Netherlands	VN	Viet Nam
CG	Congo	KE	Kenya	NO	Norway	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NZ	New Zealand	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	PL	Poland		
CM	Cameroon	KR	Republic of Korea	PT	Portugal		
CN	China	KZ	Kazakstan	RO	Romania		
CU	Cuba	LC	Saint Lucia	RU	Russian Federation		
CZ	Czech Republic	LI	Liechtenstein	SD	Sudan		
DE	Germany	LK	Sri Lanka	SE	Sweden		
DK	Denmark	LR	Liberia	SG	Singapore		
EE	Estonia						

TITLE

CAROTENOID BIOSYNTHESIS ENZYMES

This application claims the benefit of U.S. Provisional Application No. 60/083.042, filed April 24, 1998.

FIELD OF THE INVENTION

This invention is in the field of plant molecular biology. More specifically, this invention pertains to nucleic acid fragments encoding enzymes of the carotenoid biosynthesis pathway in plants and seeds.

BACKGROUND OF THE INVENTION

Plant carotenoids are orange and red lipid-soluble pigments found embedded in the membranes of chloroplasts and chromoplasts. In leaves and immature fruits the color is masked by chlorophyll but in later stages of development these pigments contribute to the bright color of flowers and fruits. Carotenoids protect against photooxidation processes and harvest light for photosynthesis. The carotenoid biosynthesis pathway leads to the production of abscisic acid with intermediaries useful in the agricultural and food industries as well as products thought to be involved in cancer prevention. (Bartley, G. E., and Scolnik, P. A. (1995) *Plant Cell* 7:1027-1038).

Phytoene desaturase transforms phytoene to zeta carotene via phytofluene. cDNAs encoding this bifunctional enzyme have been identified from bacteria, peppers, corn, Arabidopsis and Narcissus. The lightly colored zeta-carotene is converted to neurosporene by the zeta-carotene desaturase enzyme (carotene 7, 8 desaturase; EC 1.134.99.30). cDNAs encoding zeta carotene desaturase have been identified in bacteria, corn, Narcissus, tomato, Arabidopsis and pepper. Neurosporene is further desaturated into lycopene. Lycopene may have one of two different fates: through the action of lycopene epsilon cyclase it may become alpha carotene, or it may be transformed into beta carotene by lycopene cyclase. Beta-carotene dehydroxylase converts beta-carotene into zeaxanthin. Zeaxanthin epoxidase transforms zeaxanthin into violaxanthin and eventually abscisic acid.

Zeaxanthin is the bright orange product highly prized as a pigmenting agent for animal feed which makes the meat fat, skin, and egg yolks a dark yellow (Scott, M. L. et al. (1968) *Poultry Sci.* 47:863-872). Gram per gram, zeaxanthin is one of the best pigmenting compounds because it is highly absorbable. Yellow corn, which produces one of the best ratios of lutein to zeaxanthin contains in average 20 to 25 mg of xanthophyll per kg while marigold petals yield 6,000 to 10,000 mg/kg.

SUMMARY OF THE INVENTION

The instant invention relates to isolated nucleic acid fragments encoding carotenoid biosynthetic enzymes. Specifically, this invention concerns an isolated nucleic acid fragment encoding a zeta carotene desaturase or a phytoene desaturase. In addition, this invention relates to a nucleic acid fragment that is complementary to the nucleic acid fragment encoding zeta carotene desaturase or phytoene desaturase.

An additional embodiment of the instant invention pertains to a polypeptide encoding all or a substantial portion of a carotenoid biosynthetic enzyme selected from the group consisting of zeta carotene desaturase and phytoene desaturase.

In another embodiment, the instant invention relates to a chimeric gene encoding a zeta carotene desaturase or a phytoene desaturase, or to a chimeric gene that comprises a nucleic acid fragment that is complementary to a nucleic acid fragment encoding a zeta carotene desaturase or a phytoene desaturase, operably linked to suitable regulatory sequences, wherein expression of the chimeric gene results in production of levels of the encoded protein in a transformed host cell that is altered (i.e., increased or decreased) from the level produced in an untransformed host cell.

In a further embodiment, the instant invention concerns a transformed host cell comprising in its genome a chimeric gene encoding a zeta carotene desaturase or a phytoene desaturase, operably linked to suitable regulatory sequences. Expression of the chimeric gene results in production of altered levels of the encoded protein in the transformed host cell. The transformed host cell can be of eukaryotic or prokaryotic origin, and include cells derived from higher plants and microorganisms. The invention also includes transformed plants that arise from transformed host cells of higher plants, and seeds derived from such transformed plants.

An additional embodiment of the instant invention concerns a method of altering the level of expression of a zeta carotene desaturase or a phytoene desaturase in a transformed host cell comprising: a) transforming a host cell with a chimeric gene comprising a nucleic acid fragment encoding a zeta carotene desaturase or a phytoene desaturase; and b) growing the transformed host cell under conditions that are suitable for expression of the chimeric gene wherein expression of the chimeric gene results in production of altered levels of zeta carotene desaturase or phytoene desaturase in the transformed host cell.

An addition embodiment of the instant invention concerns a method for obtaining a nucleic acid fragment encoding all or a substantial portion of an amino acid sequence encoding a zeta carotene desaturase or a phytoene desaturase.

A further embodiment of the instant invention is a method for evaluating at least one compound for its ability to inhibit the activity of a zeta carotene desaturase or a phytoene desaturase, the method comprising the steps of: (a) transforming a host cell with a chimeric gene comprising a nucleic acid fragment encoding a zeta carotene desaturase or a phytoene desaturase, operably linked to suitable regulatory sequences; (b) growing the transformed host cell under conditions that are suitable for expression of the chimeric gene wherein expression of the chimeric gene results in production of zeta carotene desaturase or phytoene desaturase in the transformed host cell; (c) optionally purifying the zeta carotene desaturase or the phytoene desaturase expressed by the transformed host cell; (d) treating the zeta carotene desaturase or the phytoene desaturase with a compound to be tested; and (e) comparing the activity of the zeta carotene desaturase or the phytoene desaturase that has

been treated with a test compound to the activity of an untreated zeta carotene desaturase or phytoene desaturase, thereby selecting compounds with potential for inhibitory activity.

BRIEF DESCRIPTION OF THE DRAWINGS AND SEQUENCE DESCRIPTIONS

5 The invention can be more fully understood from the following detailed description and the accompanying drawings and Sequence Listing which form a part of this application.

Figure 1 depicts the amino acid sequence alignment between the zeta carotene desaturase from soybean clone srl.pk0022.a12 (SEQ ID NO:6), wheat clone wlm24.pk0031.e9 (SEQ ID NO:8), *Zea mays* (NCBI gi Accession No. 4105563, SEQ ID
10 NO:13) and *Capsicum annuum* (NCBI gi Accession No. 2129927, SEQ ID NO:14). Amino acids which are conserved among all sequences are indicated with an asterisk (*). Dashes are used by the program to maximize alignment of the sequences.

Figure 2 depicts the amino acid sequence alignment between the phytoene desaturase from rice clone rlr6.pk0027.d5 (SEQ ID NO:10), wheat clone wlm12.pk0003.g5 (SEQ ID
15 NO:12) and *Zea mays* (NCBI gi Accession No. 1345838, SEQ ID NO:15). Amino acids which are conserved among all sequences are indicated with an asterisk (*). Dashes are used by the program to maximize alignment of the sequences.

The following sequence descriptions and sequence listings attached hereto comply with the rules governing nucleotide and/or amino acid sequence disclosures in patent
20 applications as set forth in 37 C.F.R. §1.821-1.825.

SEQ ID NO:1 is the nucleotide sequence comprising portion of the cDNA insert in clone rls6.pk0080.a9 encoding the N-terminal fifth of a rice zeta-carotene desaturase.

SEQ ID NO:2 is the deduced amino acid sequence of the N-terminal fifth of a rice zeta-carotene desaturase derived from the nucleotide sequence of SEQ ID NO:1.

25 SEQ ID NO:3 is the nucleotide sequence comprising a portion of the cDNA insert in clone rls6.pk0080.a9 encoding the C-terminal quarter of a rice zeta-carotene desaturase.

SEQ ID NO:4 is the deduced amino acid sequence of the C-terminal quarter of a rice zeta-carotene desaturase. derived from the nucleotide sequence of SEQ ID NO:3.

30 SEQ ID NO:5 is the nucleotide sequence comprising a portion of the cDNA insert in clone srl.pk0022.a12 encoding a portion of a soybean zeta-carotene desaturase.

SEQ ID NO:6 is the deduced amino acid sequence of a portion of a soybean zeta-carotene desaturase derived from the nucleotide sequence of SEQ ID NO:5.

SEQ ID NO:7 is the nucleotide sequence comprising the entire cDNA insert in clone wlm24.pk0031.e9 encoding an entire wheat zeta-carotene desaturase.

35 SEQ ID NO:8 is the deduced amino acid sequence of an entire wheat zeta-carotene desaturase derived from the nucleotide sequence of SEQ ID NO:7.

SEQ ID NO:9 is the nucleotide sequence comprising the entire cDNA insert in clone rlr6.pk0027.d5 encoding an entire rice phytoene desaturase.

SEQ ID NO:10 is the deduced amino acid sequence of an entire rice phytoene desaturase derived from the nucleotide sequence of SEQ ID NO:9.

SEQ ID NO:11 is the nucleotide sequence comprising the entire cDNA insert in clone wlm12.pk0003.g5 encoding a nearly entire wheat phytoene desaturase.

5 SEQ ID NO:12 is the deduced amino acid sequence of a nearly entire wheat phytoene desaturase derived from the nucleotide sequence of SEQ ID NO:11.

SEQ ID NO:13 is the amino acid sequence of a *Zea mays* zeta carotene desaturase. NCBI gi Accession No. 4105563.

10 SEQ ID NO:14 is the amino acid sequence of a *Capsicum annuum* zeta carotene desaturase, NCBI gi Accession No. 2129927.

SEQ ID NO:15 is the amino acid sequence of a *Zea mays* phytoene desaturase, NCBI gi Accession No. 1345838.

15 The Sequence Listing contains the one letter code for nucleotide sequence characters and the three letter codes for amino acids as defined in conformity with the IUPAC-IUBMB standards described in *Nucleic Acids Research* 13:3021-3030 (1985) and in the *Biochemical Journal* 219 (No. 2):345-373 (1984) which are herein incorporated by reference. The symbols and format used for nucleotide and amino acid sequence data comply with the rules set forth in 37 C.F.R. §1.822.

DETAILED DESCRIPTION OF THE INVENTION

20 In the context of this disclosure, a number of terms shall be utilized. As used herein, an "isolated nucleic acid fragment" is a polymer of RNA or DNA that is single- or double-stranded, optionally containing synthetic, non-natural or altered nucleotide bases. An isolated nucleic acid fragment in the form of a polymer of DNA may be comprised of one or more segments of cDNA, genomic DNA or synthetic DNA.

25 As used herein, "substantially similar" refers to nucleic acid fragments wherein changes in one or more nucleotide bases results in substitution of one or more amino acids, but do not affect the functional properties of the protein encoded by the DNA sequence. "Substantially similar" also refers to nucleic acid fragments wherein changes in one or more nucleotide bases does not affect the ability of the nucleic acid fragment to mediate alteration of gene expression by antisense or co-suppression technology. "Substantially similar" also refers to modifications of the nucleic acid fragments of the instant invention such as deletion or insertion of one or more nucleotides that do not substantially affect the functional properties of the resulting transcript vis-à-vis the ability to mediate alteration of gene expression by antisense or co-suppression technology or alteration of the functional properties of the resulting protein molecule. It is therefore understood that the invention encompasses more than the specific exemplary sequences.

30

35

For example, it is well known in the art that antisense suppression and co-suppression of gene expression may be accomplished using nucleic acid fragments representing less than the entire coding region of a gene, and by nucleic acid fragments that do not share 100%

sequence identity with the gene to be suppressed. Moreover, alterations in a gene which result in the production of a chemically equivalent amino acid at a given site, but do not effect the functional properties of the encoded protein, are well known in the art. Thus, a codon for the amino acid alanine, a hydrophobic amino acid, may be substituted by a codon encoding another less hydrophobic residue, such as glycine, or a more hydrophobic residue, such as valine, leucine, or isoleucine. Similarly, changes which result in substitution of one negatively charged residue for another, such as aspartic acid for glutamic acid, or one positively charged residue for another, such as lysine for arginine, can also be expected to produce a functionally equivalent product. Nucleotide changes which result in alteration of the N-terminal and C-terminal portions of the protein molecule would also not be expected to alter the activity of the protein. Each of the proposed modifications is well within the routine skill in the art, as is determination of retention of biological activity of the encoded products. Moreover, the skilled artisan recognizes that substantially similar nucleic acid sequences encompassed by this invention are also defined by their ability to hybridize, under stringent conditions (0.1X SSC, 0.1% SDS, 65°C), with the sequences exemplified herein. Preferred substantially similar amino acid fragments of the instant invention are those amino acid fragments whose protein sequences are 95% identical to the protein fragments reported herein. Sequence alignments and percent similarity calculations were performed using the Megalign program of the LASARGENE bioinformatics computing suite (DNASTAR Inc., Madison, WI). Multiple alignment of the amino acid sequences was performed using the Clustal method of alignment (Higgins, D. G. and Sharp, P. M. (1989) *CABIOS*. 5:151-153) with the default parameters (GAP PENALTY=10, GAP LENGTH PENALTY=10). Default parameters for pairwise alignments using the Clustal method were KTUPLE 1, GAP PENALTY=3, WINDOW=5 and DIAGONALS SAVED=5.

A "substantial portion" of an amino acid or nucleotide sequence comprises enough of the amino acid sequence of a polypeptide or the nucleotide sequence of a gene to afford putative identification of that polypeptide or gene, either by manual evaluation of the sequence by one skilled in the art, or by computer-automated sequence comparison and identification using algorithms such as BLAST (Basic Local Alignment Search Tool; Altschul, S. F., et al., (1993) *J. Mol. Biol.* 215:403-410; see also www.ncbi.nlm.nih.gov/BLAST/). In general, a sequence of ten or more contiguous amino acids or thirty or more nucleotides is necessary in order to putatively identify a polypeptide or nucleic acid sequence as homologous to a known protein or gene. Moreover, with respect to nucleotide sequences, gene specific oligonucleotide probes comprising 20-30 contiguous nucleotides may be used in sequence-dependent methods of gene identification (e.g., Southern hybridization) and isolation (e.g., *in situ* hybridization of bacterial colonies or bacteriophage plaques). In addition, short oligonucleotides of 12-15 bases may be used as amplification primers in PCR in order to obtain a particular nucleic acid fragment comprising the primers. Accordingly, a "substantial portion" of a nucleotide sequence

comprises enough of the sequence to afford specific identification and/or isolation of a nucleic acid fragment comprising the sequence. The instant specification teaches partial or complete amino acid and nucleotide sequences encoding one or more particular plant proteins. The skilled artisan, having the benefit of the sequences as reported herein, may
5 now use all or a substantial portion of the disclosed sequences for purposes known to those skilled in this art. Accordingly, the instant invention comprises the complete sequences as reported in the accompanying Sequence Listing, as well as substantial portions of those sequences as defined above.

“Codon degeneracy” refers to divergence in the genetic code permitting variation of
10 the nucleotide sequence without effecting the amino acid sequence of an encoded polypeptide. Accordingly, the instant invention relates to any nucleic acid fragment that encodes all or a substantial portion of the amino acid sequence encoding the zeta carotene desaturase or the phytoene desaturase proteins as set forth in SEQ ID NOs:2, 4, 6, 8, 10 and 12. The skilled artisan is well aware of the “codon-bias” exhibited by a specific host cell in
15 usage of nucleotide codons to specify a given amino acid. Therefore, when synthesizing a gene for improved expression in a host cell, it is desirable to design the gene such that its frequency of codon usage approaches the frequency of preferred codon usage of the host cell.

“Synthetic genes” can be assembled from oligonucleotide building blocks that are
20 chemically synthesized using procedures known to those skilled in the art. These building blocks are ligated and annealed to form gene segments which are then enzymatically assembled to construct the entire gene. “Chemically synthesized”, as related to a sequence of DNA, means that the component nucleotides were assembled *in vitro*. Manual chemical synthesis of DNA may be accomplished using well established procedures, or automated
25 chemical synthesis can be performed using one of a number of commercially available machines. Accordingly, the genes can be tailored for optimal gene expression based on optimization of nucleotide sequence to reflect the codon bias of the host cell. The skilled artisan appreciates the likelihood of successful gene expression if codon usage is biased towards those codons favored by the host. Determination of preferred codons can be based
30 on a survey of genes derived from the host cell where sequence information is available.

“Gene” refers to a nucleic acid fragment that expresses a specific protein, including regulatory sequences preceding (5' non-coding sequences) and following (3' non-coding sequences) the coding sequence. “Native gene” refers to a gene as found in nature with its own regulatory sequences. “Chimeric gene” refers any gene that is not a native gene,
35 comprising regulatory and coding sequences that are not found together in nature. Accordingly, a chimeric gene may comprise regulatory sequences and coding sequences that are derived from different sources, or regulatory sequences and coding sequences derived from the same source, but arranged in a manner different than that found in nature.

“Endogenous gene” refers to a native gene in its natural location in the genome of an

organism. A "foreign" gene refers to a gene not normally found in the host organism, but that is introduced into the host organism by gene transfer. Foreign genes can comprise native genes inserted into a non-native organism, or chimeric genes. A "transgene" is a gene that has been introduced into the genome by a transformation procedure.

5 "Coding sequence" refers to a DNA sequence that codes for a specific amino acid sequence. "Regulatory sequences" refer to nucleotide sequences located upstream (5' non-coding sequences), within, or downstream (3' non-coding sequences) of a coding sequence, and which influence the transcription, RNA processing or stability, or translation of the associated coding sequence. Regulatory sequences may include promoters, translation
10 leader sequences, introns, and polyadenylation recognition sequences.

"Promoter" refers to a DNA sequence capable of controlling the expression of a coding sequence or functional RNA. In general, a coding sequence is located 3' to a promoter sequence. The promoter sequence consists of proximal and more distal upstream elements, the latter elements often referred to as enhancers. Accordingly, an "enhancer" is a
15 DNA sequence which can stimulate promoter activity and may be an innate element of the promoter or a heterologous element inserted to enhance the level or tissue-specificity of a promoter. Promoters may be derived in their entirety from a native gene, or be composed of different elements derived from different promoters found in nature, or even comprise synthetic DNA segments. It is understood by those skilled in the art that different promoters
20 may direct the expression of a gene in different tissues or cell types, or at different stages of development, or in response to different environmental conditions. Promoters which cause a gene to be expressed in most cell types at most times are commonly referred to as "constitutive promoters". New promoters of various types useful in plant cells are constantly being discovered; numerous examples may be found in the compilation by
25 Okamuro and Goldberg, (1989) *Biochemistry of Plants* 15:1-82. It is further recognized that since in most cases the exact boundaries of regulatory sequences have not been completely defined, DNA fragments of different lengths may have identical promoter activity.

The "translation leader sequence" refers to a DNA sequence located between the promoter sequence of a gene and the coding sequence. The translation leader sequence is
30 present in the fully processed mRNA upstream of the translation start sequence. The translation leader sequence may affect processing of the primary transcript to mRNA, mRNA stability or translation efficiency. Examples of translation leader sequences have been described (Turner, R. and Foster, G. D. (1995) *Molecular Biotechnology* 3:225).

The "3' non-coding sequences" refer to DNA sequences located downstream of a
35 coding sequence and include polyadenylation recognition sequences and other sequences encoding regulatory signals capable of affecting mRNA processing or gene expression. The polyadenylation signal is usually characterized by affecting the addition of polyadenylic acid tracts to the 3' end of the mRNA precursor. The use of different 3' non-coding sequences is exemplified by Ingelbrecht et al., (1989) *Plant Cell* 1:671-680.

“RNA transcript” refers to the product resulting from RNA polymerase-catalyzed transcription of a DNA sequence. When the RNA transcript is a perfect complementary copy of the DNA sequence, it is referred to as the primary transcript or it may be a RNA sequence derived from posttranscriptional processing of the primary transcript and is referred to as the mature RNA. “Messenger RNA (mRNA)” refers to the RNA that is without introns and that can be translated into protein by the cell. “cDNA” refers to a double-stranded DNA that is complementary to and derived from mRNA. “Sense” RNA refers to RNA transcript that includes the mRNA and so can be translated into protein by the cell. “Antisense RNA” refers to a RNA transcript that is complementary to all or part of a target primary transcript or mRNA and that blocks the expression of a target gene (U.S. Patent No. 5,107,065, incorporated herein by reference). The complementarity of an antisense RNA may be with any part of the specific gene transcript, i.e., at the 5' non-coding sequence, 3' non-coding sequence, introns, or the coding sequence. “Functional RNA” refers to sense RNA, antisense RNA, ribozyme RNA, or other RNA that may not be translated but yet has an effect on cellular processes.

The term “operably linked” refers to the association of nucleic acid sequences on a single nucleic acid fragment so that the function of one is affected by the other. For example, a promoter is operably linked with a coding sequence when it is capable of affecting the expression of that coding sequence (i.e., that the coding sequence is under the transcriptional control of the promoter). Coding sequences can be operably linked to regulatory sequences in sense or antisense orientation.

The term “expression”, as used herein, refers to the transcription and stable accumulation of sense (mRNA) or antisense RNA derived from the nucleic acid fragment of the invention. Expression may also refer to translation of mRNA into a polypeptide.

“Antisense inhibition” refers to the production of antisense RNA transcripts capable of suppressing the expression of the target protein. “Overexpression” refers to the production of a gene product in transgenic organisms that exceeds levels of production in normal or non-transformed organisms. “Co-suppression” refers to the production of sense RNA transcripts capable of suppressing the expression of identical or substantially similar foreign or endogenous genes (U.S. Patent No. 5,231,020, incorporated herein by reference).

“Altered levels” refers to the production of gene product(s) in transgenic organisms in amounts or proportions that differ from that of normal or non-transformed organisms.

“Mature” protein refers to a post-translationally processed polypeptide; i.e., one from which any pre- or propeptides present in the primary translation product have been removed.

“Precursor” protein refers to the primary product of translation of mRNA; i.e., with pre- and propeptides still present. Pre- and propeptides may be but are not limited to intracellular localization signals.

A “chloroplast transit peptide” is an amino acid sequence which is translated in conjunction with a protein and directs the protein to the chloroplast or other plastid types

present in the cell in which the protein is made. "Chloroplast transit sequence" refers to a nucleotide sequence that encodes a chloroplast transit peptide. A "signal peptide" is an amino acid sequence which is translated in conjunction with a protein and directs the protein to the secretory system (Chrispeels, J. J., (1991) *Ann. Rev. Plant Phys. Plant Mol. Biol.* 42:21-53). If the protein is to be directed to a vacuole, a vacuolar targeting signal (*supra*) can further be added, or if to the endoplasmic reticulum, an endoplasmic reticulum retention signal (*supra*) may be added. If the protein is to be directed to the nucleus, any signal peptide present should be removed and instead a nuclear localization signal included (Raikhel (1992) *Plant Phys.* 100:1627-1632).

"Transformation" refers to the transfer of a nucleic acid fragment into the genome of a host organism, resulting in genetically stable inheritance. Host organisms containing the transformed nucleic acid fragments are referred to as "transgenic" organisms. Examples of methods of plant transformation include Agrobacterium-mediated transformation (De Blaere et al. (1987) *Meth. Enzymol.* 143:277) and particle-accelerated or "gene gun" transformation technology (Klein T. M. et al. (1987) *Nature (London)* 327:70-73; U.S. Patent No. 4,945,050, incorporated herein by reference).

Standard recombinant DNA and molecular cloning techniques used herein are well known in the art and are described more fully in Sambrook, J., Fritsch, E.F. and Maniatis, T. *Molecular Cloning: A Laboratory Manual*; Cold Spring Harbor Laboratory Press: Cold Spring Harbor, 1989 (hereinafter "Maniatis").

Nucleic acid fragments encoding at least a portion of several carotenoid biosynthetic enzymes have been isolated and identified by comparison of random plant cDNA sequences to public databases containing nucleotide and protein sequences using the BLAST algorithms well known to those skilled in the art. Table 1 lists the proteins that are described herein, and the designation of the cDNA clones that comprise the nucleic acid fragments encoding these proteins.

TABLE 1
Carotenoid Biosynthetic Enzymes

Enzyme	Clone	Plant
Zeta Carotene Desaturase	rls6.pk0080.a9	Rice
	sr1.pk0022.a12	Soybean
	wlm24.pk0031.e9	Wheat
Phytoene Desaturase	rlr6.pk0027.d5	Rice
	wlm12.pk0003.g5	Wheat

The nucleic acid fragments of the instant invention may be used to isolate cDNAs and genes encoding homologous proteins from the same or other plant species. Isolation of homologous genes using sequence-dependent protocols is well known in the art. Examples

of sequence-dependent protocols include, but are not limited to, methods of nucleic acid hybridization, and methods of DNA and RNA amplification as exemplified by various uses of nucleic acid amplification technologies (e.g., polymerase chain reaction, ligase chain reaction).

5 For example, genes encoding other zeta carotene desaturases or phytoene desaturases, either as cDNAs or genomic DNAs, could be isolated directly by using all or a portion of the instant nucleic acid fragments as DNA hybridization probes to screen libraries from any desired plant employing methodology well known to those skilled in the art. Specific oligonucleotide probes based upon the instant nucleic acid sequences can be designed and
10 synthesized by methods known in the art (Maniatis). Moreover, the entire sequences can be used directly to synthesize DNA probes by methods known to the skilled artisan such as random primer DNA labeling, nick translation, or end-labeling techniques, or RNA probes using available *in vitro* transcription systems. In addition, specific primers can be designed and used to amplify a part or all of the instant sequences. The resulting amplification
15 products can be labeled directly during amplification reactions or labeled after amplification reactions, and used as probes to isolate full length cDNA or genomic fragments under conditions of appropriate stringency.

 In addition, two short segments of the instant nucleic acid fragments may be used in polymerase chain reaction protocols to amplify longer nucleic acid fragments encoding
20 homologous genes from DNA or RNA. The polymerase chain reaction may also be performed on a library of cloned nucleic acid fragments wherein the sequence of one primer is derived from the instant nucleic acid fragments, and the sequence of the other primer takes advantage of the presence of the polyadenylic acid tracts to the 3' end of the mRNA precursor encoding plant genes. Alternatively, the second primer sequence may be based
25 upon sequences derived from the cloning vector. For example, the skilled artisan can follow the RACE protocol (Frohman et al., (1988) *Proc. Natl. Acad. Sci. USA* 85:8998) to generate cDNAs by using PCR to amplify copies of the region between a single point in the transcript and the 3' or 5' end. Primers oriented in the 3' and 5' directions can be designed from the instant sequences. Using commercially available 3' RACE or 5' RACE systems
30 (BRL), specific 3' or 5' cDNA fragments can be isolated (Ohara et al., (1989) *Proc. Natl. Acad. Sci. USA* 86:5673; Loh et al., (1989) *Science* 243:217). Products generated by the 3' and 5' RACE procedures can be combined to generate full-length cDNAs (Frohman, M. A. and Martin, G. R., (1989) *Techniques* 1:165).

 Availability of the instant nucleotide and deduced amino acid sequences facilitates
35 immunological screening of cDNA expression libraries. Synthetic peptides representing portions of the instant amino acid sequences may be synthesized. These peptides can be used to immunize animals to produce polyclonal or monoclonal antibodies with specificity for peptides or proteins comprising the amino acid sequences. These antibodies can be then

be used to screen cDNA expression libraries to isolate full-length cDNA clones of interest (Lerner, R. A. (1984) *Adv. Immunol.* 36:1; Maniatis).

The nucleic acid fragments of the instant invention may be used to create transgenic plants in which the disclosed zeta carotene desaturase or phytoene desaturase are present at higher or lower levels than normal or in cell types or developmental stages in which they are not normally found. This would have the effect of altering the level of lycopene and phytoene in those cells. Manipulation of the zeta carotene desaturase or phytoene desaturase levels in transgenic plants may allow a greater accumulation of lycopene, zeaxanthin and eventually abscisic acid.

Overexpression of the zeta carotene desaturase or the phytoene desaturase proteins of the instant invention may be accomplished by first constructing a chimeric gene in which the coding region is operably linked to a promoter capable of directing expression of a gene in the desired tissues at the desired stage of development. For reasons of convenience, the chimeric gene may comprise promoter sequences and translation leader sequences derived from the same genes. 3' Non-coding sequences encoding transcription termination signals may also be provided. The instant chimeric gene may also comprise one or more introns in order to facilitate gene expression.

Plasmid vectors comprising the instant chimeric gene can then be constructed. The choice of plasmid vector is dependent upon the method that will be used to transform host plants. The skilled artisan is well aware of the genetic elements that must be present on the plasmid vector in order to successfully transform, select and propagate host cells containing the chimeric gene. The skilled artisan will also recognize that different independent transformation events will result in different levels and patterns of expression (Jones et al., (1985) *EMBO J.* 4:2411-2418; De Almeida et al., (1989) *Mol. Gen. Genetics* 218:78-86), and thus that multiple events must be screened in order to obtain lines displaying the desired expression level and pattern. Such screening may be accomplished by Southern analysis of DNA, Northern analysis of mRNA expression, Western analysis of protein expression, or phenotypic analysis.

For some applications it may be useful to direct the instant carotenoid biosynthetic enzymes to different cellular compartments, or to facilitate its secretion from the cell. It is thus envisioned that the chimeric gene described above may be further supplemented by altering the coding sequence to encode zeta carotene desaturase or phytoene desaturase with appropriate intracellular targeting sequences such as transit sequences (Keegstra, K. (1989) *Cell* 56:247-253), signal sequences or sequences encoding endoplasmic reticulum localization (Chrispeels, J. J., (1991) *Ann. Rev. Plant Phys. Plant Mol. Biol.* 42:21-53), or nuclear localization signals (Raikhel, N. (1992) *Plant Phys.* 100:1627-1632) added and/or with targeting sequences that are already present removed. While the references cited give examples of each of these, the list is not exhaustive and more targeting signals of utility may be discovered in the future.

It may also be desirable to reduce or eliminate expression of genes encoding zeta carotene desaturase or phytoene desaturase in plants for some applications. In order to accomplish this, a chimeric gene designed for co-suppression of the instant carotenoid biosynthetic enzyme can be constructed by linking a gene or gene fragment encoding a zeta carotene desaturase or a phytoene desaturase to plant promoter sequences. Alternatively, a chimeric gene designed to express antisense RNA for all or part of the instant nucleic acid fragment can be constructed by linking the gene or gene fragment in reverse orientation to plant promoter sequences. Either the co-suppression or antisense chimeric genes could be introduced into plants via transformation wherein expression of the corresponding endogenous genes are reduced or eliminated.

The instant zeta carotene desaturase or phytoene desaturase (or portions thereof) may be produced in heterologous host cells, particularly in the cells of microbial hosts, and can be used to prepare antibodies to these proteins by methods well known to those skilled in the art. The antibodies are useful for detecting zeta carotene desaturase or phytoene desaturase *in situ* in cells or *in vitro* in cell extracts. Preferred heterologous host cells for production of the instant zeta carotene desaturase or phytoene desaturase are microbial hosts. Microbial expression systems and expression vectors containing regulatory sequences that direct high level expression of foreign proteins are well known to those skilled in the art. Any of these could be used to construct a chimeric gene for production of the instant zeta carotene desaturase or phytoene desaturase. This chimeric gene could then be introduced into appropriate microorganisms via transformation to provide high level expression of the encoded carotenoid biosynthetic enzyme. An example of a vector for high level expression of the instant zeta carotene desaturase or phytoene desaturase in a bacterial host is provided (Example 7).

Additionally, the instant zeta carotene desaturase or phytoene desaturase can be used as targets to facilitate design and/or identification of inhibitors of those enzymes that may be useful as herbicides. This is desirable because the zeta carotene desaturase and the phytoene desaturase described herein catalyze various steps in carotenoid biosynthesis. Accordingly, inhibition of the activity of either one or both of the enzymes described herein could lead to inhibition plant growth. Thus, the instant zeta carotene desaturase or phytoene desaturase could be appropriate for new herbicide discovery and design.

All or a substantial portion of the nucleic acid fragments of the instant invention may also be used as probes for genetically and physically mapping the genes that they are a part of, and as markers for traits linked to those genes. Such information may be useful in plant breeding in order to develop lines with desired phenotypes. For example, the instant nucleic acid fragments may be used as restriction fragment length polymorphism (RFLP) markers. Southern blots (Maniatis) of restriction-digested plant genomic DNA may be probed with the nucleic acid fragments of the instant invention. The resulting banding patterns may then be subjected to genetic analyses using computer programs such as MapMaker (Lander et al.,

(1987) *Genomics* 1:174-181) in order to construct a genetic map. In addition, the nucleic acid fragments of the instant invention may be used to probe Southern blots containing restriction endonuclease-treated genomic DNAs of a set of individuals representing parent and progeny of a defined genetic cross. Segregation of the DNA polymorphisms is noted and used to calculate the position of the instant nucleic acid sequence in the genetic map previously obtained using this population (Botstein, D. et al., (1980) *Am. J. Hum. Genet.* 32:314-331).

The production and use of plant gene-derived probes for use in genetic mapping is described in R. Bernatzky, R. and Tanksley, S. D. (1986) *Plant Mol. Biol. Reporter* 4(1):37-41. Numerous publications describe genetic mapping of specific cDNA clones using the methodology outlined above or variations thereof. For example, F2 intercross populations, backcross populations, randomly mated populations, near isogenic lines, and other sets of individuals may be used for mapping. Such methodologies are well known to those skilled in the art.

Nucleic acid probes derived from the instant nucleic acid sequences may also be used for physical mapping (i.e., placement of sequences on physical maps; see Hoheisel, J. D., et al., In: *Nonmammalian Genomic Analysis: A Practical Guide*, Academic press 1996, pp. 319-346, and references cited therein).

In another embodiment, nucleic acid probes derived from the instant nucleic acid sequences may be used in direct fluorescence *in situ* hybridization (FISH) mapping (Trask, B. J. (1991) *Trends Genet.* 7:149-154). Although current methods of FISH mapping favor use of large clones (several to several hundred KB; see Laan, M. et al. (1995) *Genome Research* 5:13-20), improvements in sensitivity may allow performance of FISH mapping using shorter probes.

A variety of nucleic acid amplification-based methods of genetic and physical mapping may be carried out using the instant nucleic acid sequences. Examples include allele-specific amplification (Kazazian, H. H. (1989) *J. Lab. Clin. Med.* 114(2):95-96), polymorphism of PCR-amplified fragments (CAPS; Sheffield, V. C. et al. (1993) *Genomics* 16:325-332), allele-specific ligation (Landegren, U. et al. (1988) *Science* 241:1077-1080), nucleotide extension reactions (Sokolov, B. P. (1990) *Nucleic Acid Res.* 18:3671), Radiation Hybrid Mapping (Walter, M. A. et al. (1997) *Nature Genetics* 7:22-28) and Happy Mapping (Dear, P. H. and Cook, P. R. (1989) *Nucleic Acid Res.* 17:6795-6807). For these methods, the sequence of a nucleic acid fragment is used to design and produce primer pairs for use in the amplification reaction or in primer extension reactions. The design of such primers is well known to those skilled in the art. In methods employing PCR-based genetic mapping, it may be necessary to identify DNA sequence differences between the parents of the mapping cross in the region corresponding to the instant nucleic acid sequence. This, however, is generally not necessary for mapping methods.

Loss of function mutant phenotypes may be identified for the instant cDNA clones either by targeted gene disruption protocols or by identifying specific mutants for these genes contained in a maize population carrying mutations in all possible genes (Ballinger and Benzer, (1989) *Proc. Natl. Acad. Sci USA* 86:9402; Koes et al., (1995) *Proc. Natl. Acad. Sci USA* 92:8149; Bensen et al., (1995) *Plant Cell* 7:75). The latter approach may be accomplished in two ways. First, short segments of the instant nucleic acid fragments may be used in polymerase chain reaction protocols in conjunction with a mutation tag sequence primer on DNAs prepared from a population of plants in which Mutator transposons or some other mutation-causing DNA element has been introduced (see Bensen, *supra*). The amplification of a specific DNA fragment with these primers indicates the insertion of the mutation tag element in or near the plant gene encoding the zeta carotene desaturase or the phytoene desaturase. Alternatively, the instant nucleic acid fragment may be used as a hybridization probe against PCR amplification products generated from the mutation population using the mutation tag sequence primer in conjunction with an arbitrary genomic site primer, such as that for a restriction enzyme site-anchored synthetic adaptor. With either method, a plant containing a mutation in the endogenous gene encoding a zeta carotene desaturase or a phytoene desaturase can be identified and obtained. This mutant plant can then be used to determine or confirm the natural function of the zeta carotene desaturase or the phytoene desaturase gene product.

EXAMPLES

The present invention is further defined in the following Examples, in which all parts and percentages are by weight and degrees are Celsius, unless otherwise stated. It should be understood that these Examples, while indicating preferred embodiments of the invention, are given by way of illustration only. From the above discussion and these Examples, one skilled in the art can ascertain the essential characteristics of this invention, and without departing from the spirit and scope thereof, can make various changes and modifications of the invention to adapt it to various usages and conditions.

EXAMPLE 1

Composition of cDNA Libraries; Isolation and Sequencing of cDNA Clones

cDNA libraries representing mRNAs from various rice, soybean and wheat tissues were prepared. The characteristics of the libraries are described below.

TABLE 2
cDNA Libraries from Rice, Soybean and Wheat

Library	Tissue	Clone
rls6	Rice Leaf 15 Days After Germination, 6 Hours After Infection of Strain <i>Magaporthe grisea</i> 4360-R-67 (AVR2-YAMO); Susceptible	rlr6.pk0027.d5 rls6.pk0080.a9
sr1	Soybean Root	sr1.pk0022.a12
wlm12	Wheat Seedlings 12 Hours After Inoculation With <i>Erysiphe graminis f. sp tritici</i>	wlm12.pk0003.g5
wlm24	Wheat Seedlings 24 Hours After Inoculation With <i>Erysiphe graminis f. sp tritici</i>	wlm24.pk0031.e9

cDNA libraries were prepared in Uni-ZAP™ XR vectors according to the manufacturer's protocol (Stratagene Cloning Systems, La Jolla, CA). Conversion of the Uni-ZAP™ XR libraries into plasmid libraries was accomplished according to the protocol provided by Stratagene. Upon conversion, cDNA inserts were contained in the plasmid vector pBluescript. cDNA inserts from randomly picked bacterial colonies containing recombinant pBluescript plasmids were amplified via polymerase chain reaction using primers specific for vector sequences flanking the inserted cDNA sequences or plasmid DNA was prepared from cultured bacterial cells. Amplified insert DNAs or plasmid DNAs were sequenced in dye-primer sequencing reactions to generate partial cDNA sequences (expressed sequence tags or "ESTs"; see Adams, M. D. et al., (1991) *Science* 252:1651). The resulting ESTs were analyzed using a Perkin Elmer Model 377 fluorescent sequencer.

EXAMPLE 2

Identification of cDNA Clones

ESTs encoding carotenoid biosynthetic enzymes were identified by conducting BLAST (Basic Local Alignment Search Tool; Altschul, S. F., et al., (1993) *J. Mol. Biol.* 215:403-410; see also www.ncbi.nlm.nih.gov/BLAST/) searches for similarity to sequences contained in the BLAST "nr" database (comprising all non-redundant GenBank CDS translations, sequences derived from the 3-dimensional structure Brookhaven Protein Data Bank, the last major release of the SWISS-PROT protein sequence database, EMBL, and DDBJ databases). The cDNA sequences obtained in Example 1 were analyzed for similarity to all publicly available DNA sequences contained in the "nr" database using the BLASTN algorithm provided by the National Center for Biotechnology Information (NCBI). The DNA sequences were translated in all reading frames and compared for similarity to all publicly available protein sequences contained in the "nr" database using the BLASTX algorithm (Gish, W. and States, D. J. (1993) *Nature Genetics* 3:266-272) provided by the NCBI. For convenience, the P-value (probability) of observing a match of a cDNA sequence to a sequence contained in the searched databases merely by chance as calculated by BLAST are reported herein as "pLog" values, which represent the negative of the

logarithm of the reported P-value. Accordingly, the greater the pLog value, the greater the likelihood that the cDNA sequence and the BLAST "hit" represent homologous proteins.

EXAMPLE 3

Characterization of cDNA Clones Encoding Zeta Carotene Desaturase

5 The BLASTX search using the EST sequences from clones cs1n.pk0010.fl1, rls6.pk0080.a9, rlr6.pk0074.h3, srl.pk0022.a12, srl.pk0023.e7, wlm24.pk0031.e9 and wre1n.pk0058.h12 revealed similarity of the proteins encoded by the cDNAs to Zeta Carotene Desaturase from *Narcissus pseudonarcissus* and *Caspicum annum* (GenBank Accession No. AJ224683 and 1583601, respectively). Further analysis of the sequences
10 from clones rls6.pk0080.a9 and rlr6.pk0074.h3 revealed a significant region of overlap, thus affording the assembly of a contig encoding a portion of a rice Zeta Carotene Desaturase. Likewise, analysis of the sequences from clones srl.pk0022.a12 and srl.pk0023.e7 revealed a significant region of overlap, thus affording the assembly of a contig encoding a portion of a soybean Zeta Carotene Desaturase. The BLAST results for each of these ESTs and contigs
15 are shown in Table 3:

TABLE 3
BLAST Results for Clones Encoding Polypeptides Homologous
to Zeta Carotene Desaturase

Clone	Organism	GenBank Accession No.	BLAST pLog Score
cs1n.pk0010.fl1	<i>Narcissus pseudonarcissus</i>	AJ224683	50.30
Contig formed of: rls6.pk0080.a9 rlr6.pk0074.h3	<i>Caspicum annum</i>	1583601	40.30
Contig formed of srl.pk0022.a12 srl.pk0023.e7	<i>Caspicum annum</i>	1583601	10.10
wlm24.pk0031.e9	<i>Narcissus pseudonarcissus</i>	AJ224683	18.00
wre1n.pk0058.h12	<i>Narcissus pseudonarcissus</i>	AJ224683	50.70

20 The sequence of the entire cDNA insert in clone cs1n.pk0010.fl1 was determined. BLASTP analysis indicated that it is identical to zeta carotene desaturase from *Zea mays* (NCBI General Identifier No. 4105563). The 5'-terminal sequence of the cDNA insert in clone rls6.pk0080.a9 was determined and is shown in SEQ ID NO:1; this sequence includes
25 the partial sequence from clone rlr6.pk0074.h3. The deduced amino acid sequence of this cDNA is shown in SEQ ID NO:2. The 3'-terminal sequence of the cDNA insert in clone rls6.pk0080.a9 was determined and is shown in SEQ ID NO:3; the deduced amino acid sequence of this cDNA is shown in SEQ ID NO:4. The sequence of the entire cDNA insert in clone srl.pk0022.a12 was determined and is shown in SEQ ID NO:7; the deduced amino
30 acid sequence of this cDNA is shown in SEQ ID NO:8. The amino acid sequence set forth

in SEQ ID NO:8 was evaluated by BLASTP, yielding a pLog value of >254 versus the *Capsicum annum* sequence (NCBI General Identifier No. 2129927; SEQ ID NO:13). The sequence of the entire cDNA insert in clone wlm24.pk0031.e9 was determined and is shown in SEQ ID NO:7; the deduced amino acid sequence of this cDNA is shown in SEQ ID NO:8.

5 The amino acid sequence set forth in SEQ ID NO:8 was evaluated by BLASTP, yielding a pLog value of >254 versus the *Zea mays* sequence (NCBI General Identifier No. 4105563; SEQ ID NO:14). Figure 1 presents an alignment of the amino acid sequences set forth in SEQ ID NOs:6 and 8, the *Capsicum annum* sequence (SEQ ID NO:13) and the *Zea mays* sequence (SEQ ID NO:14). The amino acid sequence set forth in SEQ ID NO:6 is 82.9% similar to the *Capsicum annum* sequence and the amino acid sequence set forth in SEQ ID NO:8 is 86.4% similar to the *Zea mays* sequence.

Sequence alignments and percent similarity calculations were performed using the Megalign program of the LASARGENE bioinformatics computing suite (DNASTAR Inc., Madison, WI). Multiple alignment of the amino acid sequences was performed using the Clustal method of alignment (Higgins, D. G. and Sharp, P. M. (1989) *CABIOS*. 5:151-153) with the default parameters (GAP PENALTY=10, GAP LENGTH PENALTY=10).

Sequence alignments and BLAST scores and probabilities indicate that the instant nucleic acid fragments encode portions of a rice zeta carotene desaturase and entire soybean and wheat zeta carotene desaturases. These sequences represent the first rice, soybean and wheat sequences encoding zeta carotene desaturase.

EXAMPLE 4

Characterization of cDNA Clones Encoding Phytoene Desaturase

The BLASTX search using the EST sequences from clones rlr6.pk0027.d5 and wlm12.pk0003.g5 revealed similarity of the proteins encoded by the cDNAs to Phytoene Desaturase from *Zea maize* and *Narcissus pseudonarcissus* (GenBank Accession Nos. U37285 and X78815, respectively). The BLAST results for each of these ESTs are shown in Table 4:

TABLE 4

30 BLAST Results for Clones Encoding Polypeptides Homologous to Phytoene Desaturase

Clone	Organism	GenBank Accession No.	BLAST pLog Score
rlr6.pk0027.d5	<i>Zea maize</i>	U37285	20.15
wlm12.pk0003.g5	<i>Narcissus pseudonarcissus</i>	X78815	29.70

The sequence of the entire cDNA insert in clone rlr6.pk0027.d5 was determined and is shown in SEQ ID NO:9; the deduced amino acid sequence of this cDNA is shown in SEQ ID NO:10. The amino acid sequence set forth in SEQ ID NO:10 was evaluated by BLASTP, yielding a pLog value of >254 versus the *Zea mays* sequence (NCBI General

Identifier No. 1345838; SEQ ID NO:15). The sequence of the entire cDNA insert in clone wlm12.pk0003.g5 was determined and is shown in SEQ ID NO:11; the deduced amino acid sequence of this cDNA is shown in SEQ ID NO:12. The amino acid sequence set forth in SEQ ID NO:12 was evaluated by BLASTP, yielding a pLog value of >254 versus the *Zea mays* sequence. Figure 2 presents an alignment of the amino acid sequences set forth in SEQ ID NOs:10 and 12 and the *Zea mays* sequence (SEQ ID NO:15). The amino acid sequence set forth in SEQ ID NO:10 is 89.3% similar to the *Zea mays* sequence while the amino acid sequence set forth in SEQ ID NO:12 is 89.1% similar to the *Zea mays* sequence.

Sequence alignments and percent similarity calculations were performed using the Megalign program of the LASARGENE bioinformatics computing suite (DNASTAR Inc., Madison, WI). Multiple alignment of the amino acid sequences was performed using the Clustal method of alignment (Higgins, D. G. and Sharp, P. M. (1989) *CABIOS*. 5:151-153) with the default parameters (GAP PENALTY=10, GAP LENGTH PENALTY=10).

Sequence alignments and BLAST scores and probabilities indicate that the instant nucleic acid fragments encode entire or nearly entire rice and wheat phytoene desaturases. These sequences represent the first rice and wheat sequences encoding phytoene desaturase.

EXAMPLE 5

Expression of Chimeric Genes in Monocot Cells

A chimeric gene comprising a cDNA encoding a carotenoid biosynthetic enzyme in sense orientation with respect to the maize 27 kD zein promoter that is located 5' to the cDNA fragment, and the 10 kD zein 3' end that is located 3' to the cDNA fragment, can be constructed. The cDNA fragment of this gene may be generated by polymerase chain reaction (PCR) of the cDNA clone using appropriate oligonucleotide primers. Cloning sites (Nco I or Sma I) can be incorporated into the oligonucleotides to provide proper orientation of the DNA fragment when inserted into the digested vector pML103 as described below. Amplification is then performed in a standard PCR. The amplified DNA is then digested with restriction enzymes Nco I and SmaI and fractionated on an agarose gel. The appropriate band can be isolated from the gel and combined with a 4.9 kb Nco I-Sma I fragment of the plasmid pML103. Plasmid pML103 has been deposited under the terms of the Budapest Treaty at ATCC (American Type Culture Collection, 10801 University Blvd., Manassas, VA 20110-2209), and bears accession number ATCC 97366. The DNA segment from pML103 contains a 1.05 kb Sal I-Nco I promoter fragment of the maize 27 kD zein gene and a 0.96 kb Sma I-Sal I fragment from the 3' end of the maize 10 kD zein gene in the vector pGem9Zf(+) (Promega). Vector and insert DNA can be ligated at 15°C overnight, essentially as described (Maniatis). The ligated DNA may then be used to transform *E. coli* XL1-Blue (Epicurian Coli XL-1 Blue™; Stratagene). Bacterial transformants can be screened by restriction enzyme digestion of plasmid DNA and limited nucleotide sequence analysis using the dideoxy chain termination method (Sequenase™ DNA Sequencing Kit;

U.S. Biochemical). The resulting plasmid construct would comprise a chimeric gene encoding, in the 5' to 3' direction, the maize 27 kD zein promoter, a cDNA fragment encoding a carotenoid biosynthetic enzyme, and the 10 kD zein 3' region.

The chimeric gene described above can then be introduced into corn cells by the following procedure. Immature corn embryos can be dissected from developing caryopses derived from crosses of the inbred corn lines H99 and LH132. The embryos are isolated 10 to 11 days after pollination when they are 1.0 to 1.5 mm long. The embryos are then placed with the axis-side facing down and in contact with agarose-solidified N6 medium (Chu et al., (1975) *Sci. Sin. Peking* 18:659-668). The embryos are kept in the dark at 27°C. Friable embryogenic callus consisting of undifferentiated masses of cells with somatic proembryoids and embryoids borne on suspensor structures proliferates from the scutellum of these immature embryos. The embryogenic callus isolated from the primary explant can be cultured on N6 medium and sub-cultured on this medium every 2 to 3 weeks.

The plasmid, p35S/Ac (obtained from Dr. Peter Eckes, Hoechst Ag, Frankfurt, Germany) may be used in transformation experiments in order to provide for a selectable marker. This plasmid contains the *Pat* gene (see European Patent Publication 0 242 236) which encodes phosphinothricin acetyl transferase (PAT). The enzyme PAT confers resistance to herbicidal glutamine synthetase inhibitors such as phosphinothricin. The *pat* gene in p35S/Ac is under the control of the 35S promoter from Cauliflower Mosaic Virus (Odell et al. (1985) *Nature* 313:810-812) and the 3' region of the nopaline synthase gene from the T-DNA of the Ti plasmid of *Agrobacterium tumefaciens*.

The particle bombardment method (Klein T. M. et al., (1987) *Nature* 327:70-73) may be used to transfer genes to the callus culture cells. According to this method, gold particles (1 µm in diameter) are coated with DNA using the following technique. Ten µg of plasmid DNAs are added to 50 µL of a suspension of gold particles (60 mg per mL). Calcium chloride (50 µL of a 2.5 M solution) and spermidine free base (20 µL of a 1.0 M solution) are added to the particles. The suspension is vortexed during the addition of these solutions. After 10 minutes, the tubes are briefly centrifuged (5 sec at 15,000 rpm) and the supernatant removed. The particles are resuspended in 200 µL of absolute ethanol, centrifuged again and the supernatant removed. The ethanol rinse is performed again and the particles resuspended in a final volume of 30 µL of ethanol. An aliquot (5 µL) of the DNA-coated gold particles can be placed in the center of a Kapton™ flying disc (Bio-Rad Labs). The particles are then accelerated into the corn tissue with a Biolistic™ PDS-1000/He (Bio-Rad Instruments, Hercules CA), using a helium pressure of 1000 psi, a gap distance of 0.5 cm and a flying distance of 1.0 cm.

For bombardment, the embryogenic tissue is placed on filter paper over agarose-solidified N6 medium. The tissue is arranged as a thin lawn and covered a circular area of about 5 cm in diameter. The petri dish containing the tissue can be placed in the chamber of the PDS-1000/He approximately 8 cm from the stopping screen. The air in the chamber is

then evacuated to a vacuum of 28 inches of Hg. The macrocarrier is accelerated with a helium shock wave using a rupture membrane that bursts when the He pressure in the shock tube reaches 1000 psi.

Seven days after bombardment the tissue can be transferred to N6 medium that contains gluphosinate (2 mg per liter) and lacks casein or proline. The tissue continues to grow slowly on this medium. After an additional 2 weeks the tissue can be transferred to fresh N6 medium containing gluphosinate. After 6 weeks, areas of about 1 cm in diameter of actively growing callus can be identified on some of the plates containing the glufosinate-supplemented medium. These calli may continue to grow when sub-cultured on the selective medium.

Plants can be regenerated from the transgenic callus by first transferring clusters of tissue to N6 medium supplemented with 0.2 mg per liter of 2,4-D. After two weeks the tissue can be transferred to regeneration medium (Fromm et al., (1990) *Bio/Technology* 8:833-839).

EXAMPLE 6

Expression of Chimeric Genes in Dicot Cells

A seed-specific expression cassette composed of the promoter and transcription terminator from the gene encoding the β subunit of the seed storage protein phaseolin from the bean *Phaseolus vulgaris* (Doyle et al. (1986) *J. Biol. Chem.* 261:9228-9238) can be used for expression of the instant carotenoid biosynthetic enzyme in transformed soybean. The phaseolin cassette includes about 500 nucleotides upstream (5') from the translation initiation codon and about 1650 nucleotides downstream (3') from the translation stop codon of phaseolin. Between the 5' and 3' regions are the unique restriction endonuclease sites Nco I (which includes the ATG translation initiation codon), Sma I, Kpn I and Xba I. The entire cassette is flanked by Hind III sites.

The cDNA fragment of this gene may be generated by polymerase chain reaction (PCR) of the cDNA clone using appropriate oligonucleotide primers. Cloning sites can be incorporated into the oligonucleotides to provide proper orientation of the DNA fragment when inserted into the expression vector. Amplification is then performed as described above, and the isolated fragment is inserted into a pUC18 vector carrying the seed expression cassette.

Soybean embryos may then be transformed with the expression vector comprising sequences encoding a carotenoid biosynthetic enzyme. To induce somatic embryos, cotyledons, 3-5 mm in length dissected from surface sterilized, immature seeds of the soybean cultivar A2872, can be cultured in the light or dark at 26°C on an appropriate agar medium for 6-10 weeks. Somatic embryos which produce secondary embryos are then excised and placed into a suitable liquid medium. After repeated selection for clusters of somatic embryos which multiplied as early, globular staged embryos, the suspensions are maintained as described below.

Soybean embryogenic suspension cultures can be maintained in 35 mL liquid media on a rotary shaker, 150 rpm, at 26°C with fluorescent lights on a 16:8 hour day/night schedule. Cultures are subcultured every two weeks by inoculating approximately 35 mg of tissue into 35 mL of liquid medium.

5 Soybean embryogenic suspension cultures may then be transformed by the method of particle gun bombardment (Klein T. M. et al. (1987) *Nature* (London) 327:70-73, U.S. Patent No. 4,945,050). A DuPont Biolistic™ PDS1000/HE instrument (helium retrofit) can be used for these transformations.

10 A selectable marker gene which can be used to facilitate soybean transformation is a chimeric gene composed of the 35S promoter from Cauliflower Mosaic Virus (Odell et al. (1985) *Nature* 313:810-812), the hygromycin phosphotransferase gene from plasmid pJR225 (from *E. coli*; Gritz et al. (1983) *Gene* 25:179-188) and the 3' region of the nopaline synthase gene from the T-DNA of the Ti plasmid of *Agrobacterium tumefaciens*. The seed expression cassette comprising the phaseolin 5' region, the fragment encoding the carotenoid
15 biosynthetic enzyme and the phaseolin 3' region can be isolated as a restriction fragment. This fragment can then be inserted into a unique restriction site of the vector carrying the marker gene.

To 50 µL of a 60 mg/mL 1 µm gold particle suspension is added (in order): 5 µL DNA (1 µg/µL), 20 µl spermidine (0.1 M), and 50 µL CaCl₂ (2.5 M). The particle
20 preparation is then agitated for three minutes, spun in a microfuge for 10 seconds and the supernatant removed. The DNA-coated particles are then washed once in 400 µL 70% ethanol and resuspended in 40 µL of anhydrous ethanol. The DNA/particle suspension can be sonicated three times for one second each. Five µL of the DNA-coated gold particles are then loaded on each macro carrier disk.

25 Approximately 300-400 mg of a two-week-old suspension culture is placed in an empty 60x15 mm petri dish and the residual liquid removed from the tissue with a pipette. For each transformation experiment, approximately 5-10 plates of tissue are normally bombarded. Membrane rupture pressure is set at 1100 psi and the chamber is evacuated to a vacuum of 28 inches mercury. The tissue is placed approximately 3.5 inches away from the
30 retaining screen and bombarded three times. Following bombardment, the tissue can be divided in half and placed back into liquid and cultured as described above.

Five to seven days post bombardment, the liquid media may be exchanged with fresh media, and eleven to twelve days post bombardment with fresh media containing 50 mg/mL hygromycin. This selective media can be refreshed weekly. Seven to eight weeks post
35 bombardment, green, transformed tissue may be observed growing from untransformed, necrotic embryogenic clusters. Isolated green tissue is removed and inoculated into individual flasks to generate new, clonally propagated, transformed embryogenic suspension cultures. Each new line may be treated as an independent transformation event. These

suspensions can then be subcultured and maintained as clusters of immature embryos or regenerated into whole plants by maturation and germination of individual somatic embryos.

EXAMPLE 7

Expression of Chimeric Genes in Microbial Cells

5 The cDNAs encoding the instant carotenoid biosynthetic enzyme can be inserted into the T7 *E. coli* expression vector pBT430. This vector is a derivative of pET-3a (Rosenberg et al. (1987) *Gene* 56:125-135) which employs the bacteriophage T7 RNA polymerase/T7 promoter system. Plasmid pBT430 was constructed by first destroying the EcoR I and Hind III sites in pET-3a at their original positions. An oligonucleotide adaptor containing
10 EcoR I and Hind III sites was inserted at the BamH I site of pET-3a. This created pET-3aM with additional unique cloning sites for insertion of genes into the expression vector. Then, the Nde I site at the position of translation initiation was converted to an Nco I site using oligonucleotide-directed mutagenesis. The DNA sequence of pET-3aM in this region, 5'-CATATGG, was converted to 5'-CCCATGG in pBT430.

15 Plasmid DNA containing a cDNA may be appropriately digested to release a nucleic acid fragment encoding the protein. This fragment may then be purified on a 1% NuSieve GTG™ low melting agarose gel (FMC). Buffer and agarose contain 10 µg/ml ethidium bromide for visualization of the DNA fragment. The fragment can then be purified from the agarose gel by digestion with GELase™ (Epicentre Technologies) according to the
20 manufacturer's instructions, ethanol precipitated, dried and resuspended in 20 µL of water. Appropriate oligonucleotide adapters may be ligated to the fragment using T4 DNA ligase (New England Biolabs, Beverly, MA). The fragment containing the ligated adapters can be purified from the excess adapters using low melting agarose as described above. The vector pBT430 is digested, dephosphorylated with alkaline phosphatase (NEB) and deproteinized
25 with phenol/chloroform as described above. The prepared vector pBT430 and fragment can then be ligated at 16°C for 15 hours followed by transformation into DH5 electrocompetent cells (GIBCO BRL). Transformants can be selected on agar plates containing LB media and 100 µg/mL ampicillin. Transformants containing the gene encoding the carotenoid biosynthetic enzyme are then screened for the correct orientation with respect to the T7
30 promoter by restriction enzyme analysis.

For high level expression, a plasmid clone with the cDNA insert in the correct orientation relative to the T7 promoter can be transformed into *E. coli* strain BL21(DE3) (Studier et al. (1986) *J. Mol. Biol.* 189:113-130). Cultures are grown in LB medium containing ampicillin (100 mg/L) at 25°C. At an optical density at 600 nm of approximately
35 1, IPTG (isopropylthio-β-galactoside, the inducer) can be added to a final concentration of 0.4 mM and incubation can be continued for 3 h at 25°. Cells are then harvested by centrifugation and re-suspended in 50 µL of 50 mM Tris-HCl at pH 8.0 containing 0.1 mM DTT and 0.2 mM phenyl methylsulfonyl fluoride. A small amount of 1 mm glass beads can be added and the mixture sonicated 3 times for about 5 seconds each time with a microprobe

sonicator. The mixture is centrifuged and the protein concentration of the supernatant determined. One μg of protein from the soluble fraction of the culture can be separated by SDS-polyacrylamide gel electrophoresis. Gels can be observed for protein bands migrating at the expected molecular weight.

EXAMPLE 8

Evaluating Compounds for Their Ability to Inhibit the Activity of Carotenoid Biosynthetic Enzymes

The carotenoid biosynthetic enzymes described herein may be produced using any number of methods known to those skilled in the art. Such methods include, but are not limited to, expression in bacteria as described in Example 7, or expression in eukaryotic cell culture, *in planta*, and using viral expression systems in suitably infected organisms or cell lines. The instant carotenoid biosynthetic enzymes may be expressed either as mature forms of the proteins as observed *in vivo* or as fusion proteins by covalent attachment to a variety of enzymes, proteins or affinity tags. Common fusion protein partners include glutathione S-transferase ("GST"), thioredoxin ("Trx"), maltose binding protein, and C- and/or N-terminal hexahistidine polypeptide ("His₆"). The fusion proteins may be engineered with a protease recognition site at the fusion point so that fusion partners can be separated by protease digestion to yield intact mature enzyme. Examples of such proteases include thrombin, enterokinase and factor Xa. However, any protease can be used which specifically cleaves the peptide connecting the fusion protein and the enzyme.

Purification of the instant carotenoid biosynthetic enzymes, if desired, may utilize any number of separation technologies familiar to those skilled in the art of protein purification. Examples of such methods include, but are not limited to, homogenization, filtration, centrifugation, heat denaturation, ammonium sulfate precipitation, desalting, pH precipitation, ion exchange chromatography, hydrophobic interaction chromatography and affinity chromatography, wherein the affinity ligand represents a substrate, substrate analog or inhibitor. When the carotenoid biosynthetic enzymes are expressed as fusion proteins, the purification protocol may include the use of an affinity resin which is specific for the fusion protein tag attached to the expressed enzyme or an affinity resin containing ligands which are specific for the enzyme. For example, a carotenoid biosynthetic enzyme may be expressed as a fusion protein coupled to the C-terminus of thioredoxin. In addition, a His₆ peptide may be engineered into the N-terminus of the fused thioredoxin moiety to afford additional opportunities for affinity purification. Other suitable affinity resins could be synthesized by linking the appropriate ligands to any suitable resin such as Sepharose-4B. In an alternate embodiment, a thioredoxin fusion protein may be eluted using dithiothreitol; however, elution may be accomplished using other reagents which interact to displace the thioredoxin from the resin. These reagents include β -mercaptoethanol or other reduced thiol. The eluted fusion protein may be subjected to further purification by traditional means as stated above, if desired. Proteolytic cleavage of the thioredoxin fusion protein and the

enzyme may be accomplished after the fusion protein is purified or while the protein is still bound to the ThioBond™ affinity resin or other resin.

Crude, partially purified or purified enzyme, either alone or as a fusion protein, may be utilized in assays for the evaluation of compounds for their ability to inhibit enzymatic activation of the carotenoid biosynthetic enzymes disclosed herein. Assays may be conducted under well known experimental conditions which permit optimal enzymatic activity. For example, assays for zeta carotene desaturase are presented by Linden H. et al. (1994) *Plant Mol. Biol.* 24:369-379. Assays for phytoene desaturase are presented by Pecker I. et al. (1992) *Proc Natl Acad Sci U S A* 89:4962-4966.

10

CLAIMS

What is claimed is:

1. An isolated nucleic acid fragment encoding all or a substantial portion of a zeta carotene desaturase comprising a member selected from the group consisting of:

- 5 (a) an isolated nucleic acid fragment encoding all or a substantial portion of the amino acid sequence set forth in a member selected from the group consisting of SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6 and SEQ ID NO:8;
- 10 (b) an isolated nucleic acid fragment that is substantially similar to an isolated nucleic acid fragment encoding all or a substantial portion of the amino acid sequence set forth in a member selected from the group consisting of SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6 and SEQ ID NO:8; and
- (c) an isolated nucleic acid fragment that is complementary to (a) or (b).

15 2. The isolated nucleic acid fragment of Claim 1 wherein the nucleotide sequence of the fragment comprises all or a portion of the sequence set forth in a member selected from the group consisting of SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5 and SEQ ID NO:7.

20 3. A chimeric gene comprising the nucleic acid fragment of Claim 1 operably linked to suitable regulatory sequences.

4. A transformed host cell comprising the chimeric gene of Claim 3.

5. A zeta carotene desaturase polypeptide comprising all or a substantial portion of the amino acid sequence set forth in a member selected from the group consisting of SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6 and SEQ ID NO:8.

25 6. An isolated nucleic acid fragment encoding all or a substantial portion of a phytoene desaturase comprising a member selected from the group consisting of:

- 30 (a) an isolated nucleic acid fragment encoding all or a substantial portion of the amino acid sequence set forth in a member selected from the group consisting of SEQ ID NO:10 and SEQ ID NO:12;
- (b) an isolated nucleic acid fragment that is substantially similar to an isolated nucleic acid fragment encoding all or a substantial portion of the amino acid sequence set forth in a member selected from the group consisting of SEQ ID NO:10 and SEQ ID NO:12; and
- (c) an isolated nucleic acid fragment that is complementary to (a) or (b).

35 7. The isolated nucleic acid fragment of Claim 6 wherein the nucleotide sequence of the fragment comprises all or a portion of the sequence set forth in a member selected from the group consisting of SEQ ID NO:9 and SEQ ID NO:11.

8. A chimeric gene comprising the nucleic acid fragment of Claim 6 operably linked to suitable regulatory sequences.

9. A transformed host cell comprising the chimeric gene of Claim 8.

10. A phytoene desaturase polypeptide comprising all or a substantial portion of the amino acid sequence set forth in a member selected from the group consisting of SEQ ID NO:10 and SEQ ID NO:12.

5 11. A method of altering the level of expression of a carotenoid biosynthetic enzyme in a host cell comprising:

(a) transforming a host cell with the chimeric gene of any of Claims 3 and 8; and

10 (b) growing the transformed host cell produced in step (a) under conditions that are suitable for expression of the chimeric gene wherein expression of the chimeric gene results in production of altered levels of a carotenoid biosynthetic enzyme in the transformed host cell.

12. A method of obtaining a nucleic acid fragment encoding all or a substantial portion of the amino acid sequence encoding a carotenoid biosynthetic enzyme comprising:

15 (a) probing a cDNA or genomic library with the nucleic acid fragment of any of Claims 1 and 6;

(b) identifying a DNA clone that hybridizes with the nucleic acid fragment of any of Claims 1 and 6;

(c) isolating the DNA clone identified in step (b); and

20 (d) sequencing the cDNA or genomic fragment that comprises the clone isolated in step (c)

wherein the sequenced nucleic acid fragment encodes all or a substantial portion of the amino acid sequence encoding a carotenoid biosynthetic enzyme.

25 13. A method of obtaining a nucleic acid fragment encoding a substantial portion of an amino acid sequence encoding a carotenoid biosynthetic enzyme comprising:

(a) synthesizing an oligonucleotide primer corresponding to a portion of the sequence set forth in any of SEQ ID NOs:1, 3, 5, 7, 9 and 11; and

30 (b) amplifying a cDNA insert present in a cloning vector using the oligonucleotide primer of step (a) and a primer representing sequences of the cloning vector

wherein the amplified nucleic acid fragment encodes a substantial portion of an amino acid sequence encoding a carotenoid biosynthetic enzyme.

14. The product of the method of Claim 12.

15. The product of the method of Claim 13.

35 16. A method for evaluating at least one compound for its ability to inhibit the activity of a carotenoid biosynthetic enzyme, the method comprising the steps of:

(a) transforming a host cell with a chimeric gene comprising a nucleic acid fragment encoding a carotenoid biosynthetic enzyme, operably linked to suitable regulatory sequences;

- 5 (b) growing the transformed host cell under conditions that are suitable for expression of the chimeric gene wherein expression of the chimeric gene results in production of the carotenoid biosynthetic enzyme encoded by the operably linked nucleic acid fragment in the transformed host cell;
- (c) optionally purifying the carotenoid biosynthetic enzyme expressed by the transformed host cell;
- (d) treating the carotenoid biosynthetic enzyme with a compound to be tested; and
- 10 (e) comparing the activity of the carotenoid biosynthetic enzyme that has been treated with a test compound to the activity of an untreated carotenoid biosynthetic enzyme,
- thereby selecting compounds with potential for inhibitory activity.

FIGURE 1

```

**          *          *
SEQ ID NO:13 MASVAA--TTTLAPALAPRRARPGTG-----LVPPR--RASAVAARSTVTSPTWR-
SEQ ID NO:14 MATCSAYLCCPATASLKKRVFPDGSAGLFFGRRRLSNRLV-TPKSVIRADLNSMVSDM
SEQ ID NO:06 MAS---LIQCSATS-----LSAVPITTRFTRTHKSRLRCSLDANVSDM
SEQ ID NO:08 MA--AT--SCALVSALVVGRRR---G-----PSCQR--AAAAGVVRCSLDSKVSDM
1
*****
--QRSQRLFPPEPEHYRGPCLKVAI IGAGLAGMSTAVELLDQGHEVDLYESRPFIGGKVG
STNAPKGLFPPEPEHYRGPCLKVAI IGAGLAGMSTAVELLDQGHEVDIYESRTFIGGKVG
SVNAPKGLFPPEPEHYRGPCLKVAI IGAGLAGMSTAVELLDQGHEVDIYESRPFIGGKVG
AINAPKGLFPPEPEHYRGPCLKVAI IGAGLAGMSTAVELLDQGHEVDLYDSRTFIGGKVG
61
*****
***          *          *          *          *          *          *          *
SFVDRQGNHIEMGLHVFFGCGYNNLFRMLMKKVGADNNLLVKEHTHTFVNKGGTIGELDFRF
SFVDKRGNHIEMGLHVFFGCGYNNLFRMLMKKVGAEKNLLVKEHTHTFVNKGGIIGELDFRF
SFVDKGGNHIEMGLHVFFGCGYNNLFRLLKKVGAENLLVKDHTHTFVNKGGQIGELDFRF
SFVDKHGNHIEMGLHVFFGCGYNNLFRMLMKKVGADNNLLVKEHTHTFVNKGGIVGELDFRF
121
*****
PVGAPLHGIQAFRLTNQLKVYDKARNAVALALSPVVRRALVDPDGAQQVRDLDDISFSDW
PVGAPLHGINAFLSTNQLKTYDKARNAVALALSPVVRRALVDPDGAQQIRDLDSVSFSDW
PIGAPIHGIRAFLLTNQLNTYDKARNAVALALSPVVRRALVDPDGAQRDIRNLDISFSDW
PVGAPLHGIQAFRLTNQLKVYDKARNAVALALSPVVRRALVDPDGAQQVRDLDDVSFTDW
181

```

FIGURE 1 - CONTINUED

```

* * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * *
SEQ ID NO:13 FMSKGGTRESITRMWDPVRYALGFIDCDNISARCMLTI FTLFATKTEASLLRLKGS PDV
SEQ ID NO:14 FMSKGGTRASIORMWDPVAYALGFIDCDNISARCMLTI FALFATKTEASLLRLKGS PDV
SEQ ID NO:06 FLSKGGTRMSITKMWDPVAYALGFIDCDNISARCMLTI FALFATKTEASLLRLKGS PDV
SEQ ID NO:08 FMSRGGTRESITRMWDPVAYALGFIDCDNISARCMLTI FTLFATKTEASLLRLKGS PDV
241
* * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * *
SEQ ID NO:13 YLSGPIKKYITDRGGRFHLRWGCREVLYEKS PDGETYVKGLLLTKATSREIIKADAYVAA
SEQ ID NO:14 YLSGPIKKYIIDKGGRFHLRWGCREVLYETSSDGS MYVSGLAMSKATQKKI VKADAYVAA
SEQ ID NO:06 YLSGPIRKYIMDRGGRFHLRWGCRELLYDKSADGSIYVTGLSMSKATAKKI VKADAYVAA
SEQ ID NO:08 YLSGPIKKYITDRGGRFHLRWGCREVLYDKSPDGETYVKGFLISKATSSEIIKADAYVAA
301
* * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * *
SEQ ID NO:13 CDVPGIKRLLPSEWREWEFMDNIYKLDGVPVTVQLRYNGWVTELDLEKSRQLQRAVGL
SEQ ID NO:14 CVVPGIKRLLVPQKWRELEFFGNIYKLIGVPVTVQLRYNGWVTELDLERSRQSKRATGL
SEQ ID NO:06 CDVPGIKRLLPSEWREQEFFFNNIYELVGVPVTVQLRYNGWVTELDLEKSRRLGKAVGL
SEQ ID NO:08 CDVPGIKRLLPSEWREWDMFDNIYKLDGVPVTVQLRYNGWVTEVQDLEKSRQLQKAVGL
361
* * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * *
SEQ ID NO:13 DNLLYTADADFSCFSDALSSPADYYIEGQGLIQAVLTPGDPYMLPNEEIIISKVQKQV
SEQ ID NO:14 DNLLYTPDADFSCFADLALASPEDYYIEGQGLLCVLTGDPYMLPNEEIIIRRVSKQV
SEQ ID NO:06 DNLLYTPDADFSCFADLALSSPEDYYIEGQGLLCVLTGDPYMLPND EIIARVAKQV
SEQ ID NO:08 DNLLYTPDADFSCFSDALSSPADYYIEGQGLIQAVLTPGDPYMLPNEEIIISKVEKQV
421

```

FIGURE 1 - CONTINUED

```

***          *****          ***          *****          *****
SEQ ID NO:13  VELFPSSRGLEVTWSSVVKIGQSLYREAPGNDPFRPDQKTPVKNFFLSGSYTKQDYIDSM
SEQ ID NO:14  LALFPSSQGLEVTWSSVVKIGQSLYREGPGKDPFRPDQKTPVENFFLAGSYTKQDYIDSM
SEQ ID NO:06  LALFPSSQGLEVTWSSVVKIGQSLYREGPGKDPYRPDQKTPVRNFFLAGSYTKQDYIDSM
SEQ ID NO:08  LDLFPSARGLEVTWSSVVKIGQSLYREAPGNDPFRPDQKTPVKNFFLSGSYTKQDYIDSM
481
540

```

SEQ ID NO: 13	EGATLSGRRTSAYICGAGEELLALRKKLL---	IDDGEKALGNVQVLQAS	*
SEQ ID NO: 14	EGATLSGRQASAYICDAGEQLLALRKKIAAAELNEISKGVSLSDELSLV		
SEQ ID NO: 06	EGATLSGRQASAYICDAGEELVALRKKLDA-EFKDDLKISNTKDELSLV		
SEQ ID NO: 08	EGATLSGRRTAAYICGAGEELLAIRKKLI---	VDHSEKASGMVQMLQTS	
	541	589	

FIGURE 2

SEQ ID NO:15	MDTGCLSSMNI	TGASQTRSFAGQLPPQRCFASS	-----HYTSFAVKKLVSRNKGRRS
SEQ ID NO:10	MDTGCLSSMNI	TGTSQARSFAGQLPTHR	CFASSIQALKSSQHVSGVKS
SEQ ID NO:12		-----	-----
			60
			1

SEQ ID NO:15	HRRHPALQVVC	KDFPRP	LESTINYLEAGQLSSFFRN	SERPSKPLQVVAGAGLAGLSTA
SEQ ID NO:10	RRRLGALQVVC	KDFPRP	LENTINFLEAGQLSSFFRN	SEQPTKPLQVVIAGAGLAGLSTA
SEQ ID NO:12		-----	-----	-----
				120
				61

SEQ ID NO:15	KYLADAGHKPIL	LEARDVLGGKVA	AWKDE	DGDWYETGLHIFFGAYPNIQNLFGELRIEDR	* * * * *
SEQ ID NO:10	KYLADAGHKPIL	LEARDVLGGKIA	AWKDE	DGDWYETGLHIFFGAYPNIQNLFGELGINDR	
SEQ ID NO:12		-----	-----	-----TRFAELGISDR	
					180
					121

SEQ ID NO:15	LQWKEHSMIFAM	PNKPGEFSR	FDPE	ETLPAPINGI	WAILRNNEMLTWPEKVKFAIGLLPA	*****
SEQ ID NO:10	LQWKEHSMIFAM	PNKPGEFSR	FDPE	ETLPAPLNGI	WAILRNNEMLTWPEKVKFAIGLLPA	
SEQ ID NO:12	LQWKEHSMIFAM	PNKPGEYSR	FDPE	ETLPAPLNGV	WAILKNNEMLTWPEKVKFAIGLLPA	
						240
						181

FIGURE 2 -CONTINUED

SEQ ID NO:15	MVGGPYVEAQDGLTVSEWMKKQGVDPDRVNDVFIAMSKALNFINPDELSMQCILIALNR	241
SEQ ID NO:10	MVGGAAYVEAQDGLTVSEWMKKQGVDPDRVNDVFIAMSKALNFINPDELSMQCILIALNR	300
SEQ ID NO:12	MLGGQAYVEAQDGLTVSEWMEKQGVDPDRVNDVFIAMSKALNFINPDELSMQCILIALNR	300

SEQ ID NO:15	FLQEKHGSKMAFLDGNPPPERLCMPIVDHIRSRGGEVRLNSRIKKIELNPDGTVKHFALSD	301
SEQ ID NO:10	FLQEKHGSKMAFLDGNPPPERLCMPIVDHVRSLGGEVRLNSRIQKIELNPDGTVKHFALTD	360
SEQ ID NO:12	FLQEKHGSKMAFLDGNPPPERLCMPIVNHISLGGGEVRLNSRIQKIELNPDGTVKHFALTD	360

SEQ ID NO:15	GTQITGDAYVCATPVDIFKLLVPQEWSEITYFKKLEKLVGVPVINVHIWFDRLKLNNTYDH	420
SEQ ID NO:10	GTQITGDAYVFATPVDILKLLVPQEWKEISYFKKLEKLVGVPVINVHIWFDRLKLNNTYDH	420
SEQ ID NO:12	GTQITGDAYVFAAPVDIFKLLVPQEWREISYFKRLDKLVGVPVINVHIWFDRLKLNNTYDH	420

SEQ ID NO:15	LLFSRSSLLSVYADMSVTCKEYYDPPNRSMLELVFAPADEWIGRSDTEIIDATMEELAKLF	480
SEQ ID NO:10	LLFSRSSLLSVYADMSVTCKEYYDQKRSMLELVFAPAEWVGRSDTEIEATMQELAKLF	480
SEQ ID NO:12	LLFSRSSLLSVYADMSLACKKEYYDPPNRSMLELVFAPAEWIGRSDTEIEATMLELAKLF	480

FIGURE 2 - CONTINUED

SEQ ID NO:15	PDEIAADQSKAKILKYHIVKTPRSVYKTVPNCEPCRP LQ RSP IEGFYLAGDYTKQKYLAS	*****
SEQ ID NO:10	PDEIAADQSKAKILKYHVVKTPRSVYKTI PDCEPCRP LQ RSP IEGFYLAGDYTKQKYLAS	*****
SEQ ID NO:12	PDEIAADQSKAKILKYHVVKTPRSVYKTVPNCEPCRP LQ RSP IEGFYLAGDYTKQKYLAS	*****
		540

481

SEQ ID NO:15	MEGAVLSGKLCAQSI VQDYSRLALRSQKSLQSGEVPVP---S	*****
SEQ ID NO:10	MEGAVLSGKLCAQSVVEDYKMLSRRLKSLQS-EVPVAS.--	*****
SEQ ID NO:12	MEGAVLSGKFCAQSI VQDSKMLSRRLS QESLQS-EAPVASKL.	*****
		582

541

SEQUENCE LISTING

<110> E. I. DU PONT DE NEMOURS AND COMPANY

<120> CAROTENOID BIOSYNTHESIS ENZYMES

<130> BB-1115-A

<140>

<141>

<150> 60/083,042

<151> APRIL 24, 1998

<160> 15

<170> MICROSOFT OFFICE 97

<210> 1

<211> 552

<212> DNA

<213> Oryza sativa

<400> 1

```

gcacgaggtt ctaaccagtg gtgcgcgcgt ggtggcgcca cggaggccga caaatccgct 60
gctcgacggc atcttcgcct cctccccac cgcctcctc ttccacctct cgcgctcacg 120
ccatttatac tctcctcctc ctctcctcc tctactccta gcagcagcgt tggttggtga 180
gctgaggcat atggccatgg ctgccacgtc cgcgcgcgcg tccacgctgg ccccggcgctc 240
gttctccgcc gccggcggca gcaggaggcg gcgcgggtgc ccgaatcca ggggtgcgggt 300
gggggtgggg gtgcggtgct cgctcgacag caacgtctcc gacatggccg tcaacgctcc 360
gaaaggattg ttcccgccgg agccggagca ctacagggg ccgaagctga aggtggccat 420
catatgggcc gggctcgccg gcatgtccac cgcggtggag ctcttgacc agtgccatga 480
ggttgatctg tacgagtcct ggccgtttat cgggtggaaa gtcggttctt ttgtggatac 540
gaaaggcaac ag                                     552

```

<210> 2

<211> 120

<212> PRT

<213> Oryza sativa

<400> 2

```

Met Ala Met Ala Ala Thr Ser Arg Ala Pro Ser Thr Leu Ala Pro Ala
 1              5              10              15

```

```

Ser Phe Ser Ala Ala Gly Gly Ser Arg Arg Arg Arg Arg Cys Pro Asn
          20              25              30

```

```

Pro Arg Val Arg Val Gly Val Gly Val Arg Cys Ser Leu Asp Ser Asn
          35              40              45

```

```

Val Ser Asp Met Ala Val Asn Ala Pro Lys Gly Leu Phe Pro Pro Glu
          50              55              60

```

```

Pro Glu His Tyr Arg Gly Pro Lys Leu Lys Val Ala Ile Ile Trp Ala
          65              70              75              80

```

```

Gly Leu Ala Gly Met Ser Thr Ala Val Glu Leu Leu Asp Gln Cys His
          85              90              95

```

```

Glu Val Asp Leu Tyr Glu Ser Arg Pro Phe Ile Gly Gly Lys Val Gly
          100              105              110

```

Ser Phe Val Asp Thr Lys Gly Asn
115 120

<210> 3
<211> 844
<212> DNA
<213> Oryza sativa

<400> 3
 tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 60
 tttttttttt tttttaagca ccatctctca caagatttta agtgatgtgt gcactatata 120
 aacatcatca tcatcatcaa tgggttcaat gatcgggttaa gctgaacaaa acacacgctg 180
 cctgatgtaa tcccgtatgc cacctttggt tcccaatcac aaattcacia taacgaagaa 240
 cacaccattg acgggcgaac gaacgatcgt gtcaaggcaa aaagggcagc agcagattgg 300
 acctaaatgc tcaatcaagt aggatttaca tgagagtcgc tgcggaaatc gcctactaca 360
 cccaggaagc tgtttttggt ttgggcatag caatccactc actccagcat gtgtcatttg 420
 aggaagctca gcttgtctga agggcctcga ccttaccctt ggccttttcg ctgtcgtcga 480
 caatgagctt ctttcggagg gcgagcagct cctctcctgc accacagatg taggcgcgcg 540
 ttctcctgcc tgagagagtt gcccttccca tgetgtcaat gtagtcctgt tttgtgtaag 600
 agccagacag gaagaagttt ttaactgggt tcttttgatc aggtctaaat ggattaattc 660
 ctggggactc gcggtacaat gattgaccga ttttcaccaa actcgaccaa gttagttcca 720
 agccctggga tgacgggaac aattcctaga cctgggtttg gaccctggca aataaccctt 780
 caatccggaa tgggaagtta aggatgccaa gggttaacac aacttggtt aaagaacctt 840
 gcc 844

<210> 4
<211> 138
<212> PRT
<213> Oryza sativa

<220>
<221> UNSURE
<222> (33)

<400> 4
 Gly Gln Gly Ser Leu Ile Gln Val Val Leu Thr Leu Gly Ile Leu Asn
1 5 10 15
 Phe Pro Phe Arg Ile Glu Gly Leu Phe Ala Arg Val Gln Asn Gln Val
20 25 30
 Xaa Glu Leu Phe Pro Ser Ser Gln Gly Leu Glu Leu Thr Trp Ser Ser
35 40 45
 Leu Val Lys Ile Gly Gln Ser Leu Tyr Arg Glu Ser Pro Gly Ile Asn
50 55 60
 Pro Phe Arg Pro Asp Gln Lys Thr Pro Val Lys Asn Phe Phe Leu Ser
65 70 75 80
 Gly Ser Tyr Thr Lys Gln Asp Tyr Ile Asp Ser Met Glu Gly Ala Thr
85 90 95
 Leu Ser Gly Arg Arg Thr Ala Ala Tyr Ile Cys Gly Ala Gly Glu Glu
100 105 110
 Leu Leu Ala Leu Arg Lys Lys Leu Ile Val Asp Asp Ser Glu Lys Ala
115 120 125

Arg Gly Lys Val Glu Ala Leu Gln Thr Ser
130 135

<210> 5
<211> 1867
<212> DNA
<213> Glycine max

<400> 5
gcacgagctt tcaccttcac ctatggccat ggcttctttg attcagtgtt ctgcaacttc 60
cctctccgct gtcccatca ctacccgttt caccaggacc cacaagtctc gccttcgctg 120
ctccttagat gctaattgtt ccgacatgag cgtaaacgag ccaaaagggt tgtttcctcc 180
ggaacctgaa cattatcgag gacaaaaact gaaagtggca attattggag ctggacttgc 240
aggcatgtca actgcagtgg aactcttgga tcaaggccat gaggtggata tctatgagtc 300
aagacctttt attggtggca aagttggctc ttttgttgac aaagggtggga atcacattga 360
aatgggattg catgttttct ttggttgcta caacaatctt ttccgattgt tgaagaagggt 420
gggtgcagaa aacaatctac ttgtgaagga tcacactcac acttttgta acaaaggggg 480
tcaaattggt gaactggatt ttcgcttccc aattggggca ccaatacacg ggataagggc 540
gtttttgacc acaaatcagc ttaataactta tgataaggct agaaatgctg tggctcttgc 600
actcagtcga gttgtcagag ctcttggtga tccagatggg gcactgaggg acataaggaa 660
tttgatagat attagctttt cagattgggt tttatccaaa ggtggcacac gcatgagtat 720
tacaaaaatg tgggatccag ttgcctatgc ccttgggttt atcgactgtg ataatatcag 780
tgctcgtgac atgctcacca tatttgcatt gtttgccaca aagaccgagg ctccctttt 840
gcgaatgctg aagggttcac cggatgttta tctgagtggc cccatccgaa agtacatcat 900
ggacagaggg ggcaggttcc atcttaggtg gggatgcaga gaactgcttt atgacaaatc 960
tgctgatggg agtatattatg ttacaggact ttccatgtca aaggccactg ccaagaaaat 1020
tgtgaaagct gatgcttatg ttgctgcttg tgatgtccct ggaattaaga gattacttcc 1080
atcagagtgg agggaacagg agtttttcaa taatatctat gaactagtgt gagttcctgt 1140
agtcacagtg caactcagat acaatgggtg gggttacagag ttgcaggatc tagaaaagtc 1200
aaggcgactg gggaaagctg ttgggttgga taacctctc tatacacccg atgcagattt 1260
ttcttgcttt gcagacctg cactttcatc tccagaggat tattacattg agggacaagg 1320
atcattgctc caatgtgttc tgacgccagg agatccatac atgcccttac caaatgacga 1380
aattattgca agggtagcaa aacaggtttt ggcactgttc ccatcatctc aagggttgga 1440
agtgacttgg tcttctgttg ttaaaattgg ccaatctctg taccgtgagg gtcctggtaa 1500
agatccatat agacctgac aaaagacacc agtgaggaat ttctttcttg ctggctctta 1560
cacaaagcag gactacatag acagcatgga aggtgcaacc ttgtctggca gacaagcttc 1620
ggcctatata tgcgatgcag gggaagaatt agtggcattg cggaagaagc tcgacgctga 1680
gtttaaagac gacttaaaaa tatcaaatc taaagatgaa ttgagcttag tatgataaat 1740
tttcatgagg tggcatgtt gacatgaagc aatctgggtt ttttgcattg aacatttttc 1800
aatgtatatt caatcgataa ataattacag aatagattgt ttcgaagcaa aaaaaaaaaa 1860
aaaaaaaa 1867

<210> 6
<211> 568
<212> PRT
<213> Glycine max

<400> 6
Met Ala Ser Leu Ile Gln Cys Ser Ala Thr Ser Leu Ser Ala Val Pro
1 5 10 15
Ile Thr Thr Arg Phe Thr Arg Thr His Lys Ser Arg Leu Arg Cys Ser
20 25 30
Leu Asp Ala Asn Val Ser Asp Met Ser Val Asn Ala Pro Lys Gly Leu
35 40 45
Phe Pro Pro Glu Pro Glu His Tyr Arg Gly Pro Lys Leu Lys Val Ala
50 55 60

Ile Ile Gly Ala Gly Leu Ala Gly Met Ser Thr Ala Val Glu Leu Leu
 65 70 75 80
 Asp Gln Gly His Glu Val Asp Ile Tyr Glu Ser Arg Pro Phe Ile Gly
 85 90 95
 Gly Lys Val Gly Ser Phe Val Asp Lys Gly Gly Asn His Ile Glu Met
 100 105 110
 Gly Leu His Val Phe Phe Gly Cys Tyr Asn Asn Leu Phe Arg Leu Leu
 115 120 125
 Lys Lys Val Gly Ala Glu Asn Asn Leu Leu Val Lys Asp His Thr His
 130 135 140
 Thr Phe Val Asn Lys Gly Gly Gln Ile Gly Glu Leu Asp Phe Arg Phe
 145 150 155 160
 Pro Ile Gly Ala Pro Ile His Gly Ile Arg Ala Phe Leu Thr Thr Asn
 165 170 175
 Gln Leu Asn Thr Tyr Asp Lys Ala Arg Asn Ala Val Ala Leu Ala Leu
 180 185 190
 Ser Pro Val Val Arg Ala Leu Val Asp Pro Asp Gly Ala Leu Arg Asp
 195 200 205
 Ile Arg Asn Leu Asp Ser Ile Ser Phe Ser Asp Trp Phe Leu Ser Lys
 210 215 220
 Gly Gly Thr Arg Met Ser Ile Thr Lys Met Trp Asp Pro Val Ala Tyr
 225 230 235 240
 Ala Leu Gly Phe Ile Asp Cys Asp Asn Ile Ser Ala Arg Cys Met Leu
 245 250 255
 Thr Ile Phe Ala Leu Phe Ala Thr Lys Thr Glu Ala Ser Leu Leu Arg
 260 265 270
 Met Leu Lys Gly Ser Pro Asp Val Tyr Leu Ser Gly Pro Ile Arg Lys
 275 280 285
 Tyr Ile Met Asp Arg Gly Gly Arg Phe His Leu Arg Trp Gly Cys Arg
 290 295 300
 Glu Leu Leu Tyr Asp Lys Ser Ala Asp Gly Ser Ile Tyr Val Thr Gly
 305 310 315 320
 Leu Ser Met Ser Lys Ala Thr Ala Lys Lys Ile Val Lys Ala Asp Ala
 325 330 335
 Tyr Val Ala Ala Cys Asp Val Pro Gly Ile Lys Arg Leu Leu Pro Ser
 340 345 350
 Glu Trp Arg Glu Gln Glu Phe Phe Asn Asn Ile Tyr Glu Leu Val Gly
 355 360 365
 Val Pro Val Val Thr Val Gln Leu Arg Tyr Asn Gly Trp Val Thr Glu
 370 375 380

Leu Gln Asp Leu Glu Lys Ser Arg Arg Leu Gly Lys Ala Val Gly Leu
 385 390 395 400
 Asp Asn Leu Leu Tyr Thr Pro Asp Ala Asp Phe Ser Cys Phe Ala Asp
 405 410 415
 Leu Ala Leu Ser Ser Pro Glu Asp Tyr Tyr Ile Glu Gly Gln Gly Ser
 420 425 430
 Leu Leu Gln Cys Val Leu Thr Pro Gly Asp Pro Tyr Met Pro Leu Pro
 435 440 445
 Asn Asp Glu Ile Ile Ala Arg Val Ala Lys Gln Val Leu Ala Leu Phe
 450 455 460
 Pro Ser Ser Gln Gly Leu Glu Val Thr Trp Ser Ser Val Val Lys Ile
 465 470 475 480
 Gly Gln Ser Leu Tyr Arg Glu Gly Pro Gly Lys Asp Pro Tyr Arg Pro
 485 490 495
 Asp Gln Lys Thr Pro Val Arg Asn Phe Phe Leu Ala Gly Ser Tyr Thr
 500 505 510
 Lys Gln Asp Tyr Ile Asp Ser Met Glu Gly Ala Thr Leu Ser Gly Arg
 515 520 525
 Gln Ala Ser Ala Tyr Ile Cys Asp Ala Gly Glu Glu Leu Val Ala Leu
 530 535 540
 Arg Lys Lys Leu Asp Ala Glu Phe Lys Asp Asp Leu Lys Ile Ser Asn
 545 550 555 560
 Thr Lys Asp Glu Leu Ser Leu Val
 565

<210> 7
 <211> 2074
 <212> DNA
 <213> Triticum aestivum

<400> 7
 cgctcacaca atttatgcta ttcctcgccc agtggctcca tcttcccggc tccgtccctc 60
 actgctcctc gggttttgta ctgcttttgc atcgacgagg ccacggccat ggccgcccacg 120
 tcgtgcgcgc tagtctcggc cctcgtagtc ggccggcgcc gcgggcccgc ctgccagcgc 180
 gcggcgccgc ccggtgtggt tcgatgctcg ctgcacagca aggtctccga catggccatc 240
 aacgcaccga aaggactgtt ccgcgcggag cctgagcact acaggggccc gaagctcaag 300
 gtcgccatca taggtgccgg cctcgccggc atgtccaccg cagtagagct cttggaccag 360
 ggacatgagg ttgatctgta tgactcccga acttttattg gcggcaaggt tggttctttt 420
 gtcgacaagc atgggaacca tatcgagatg gggctgcatg tcttctttgg ttgttacagc 480
 aatcttttcc gcctcatgaa gaaggttggg gctgataata atctactagt caaggaacat 540
 acccatactt ttgtaaataa agggggcatt gttggtgaac ttgatttttcg gttccctgtg 600
 ggagctccat tacatggtat ccaagcattt ctaagaacca atcaactcaa ggtttatgat 660
 aaggcaagga atgcagttgc tctcgcccta agcccagttg ttcgagctct tcttgatcca 720
 gacggtgcat tgcaacaagt acgggacttg gatgatgtaa gtttactga ttggttcacg 780
 tctagaggtg gtactcgaga gagcatcaca agaattgtgg atcctgttgc ttatgctctt 840
 ggttttatcg actgtgataa tatcagtgtc cgatgcatgc ttactatttt caccctgttt 900
 gccacaaaga cagaggcatc tttgttgctc atgctaaagg gctcacctga tgtttactta 960
 agtggcccaa taaagaagta cataacagac aggggttgta ggtttcactt gaagtgggga 1020
 tgccgagagg ttctctacga taagtcaccc gatggagaga cctatgtgaa aggctttctc 1080
 atctctaagg ctacaagtag tgagataatc aaagcagatg catatgttgc agcttgtgat 1140


```

gtcccagggg tcaaaaagact attaccatcg gaatggaggg aatgggatat gtttgacaat 1200
atatacaagt tagatgggtgt tcctgtagtc actgttcagc ttcgctacaa tggatggggtt 1260
actgaagtcc aagacttgga gaaatcaaga caactccaaa aagcagttgg cttgggataat 1320
ctactctata ctccagatgc agacttctcc tgtttttcag accttgcaact gtcactctcct 1380
gctgactact acattgaagg acaagggttcc ctgatccaag ctgtgctaac tccagggtgat 1440
ccgtacatgc cattgccaaa tgaggagatt attagcaagg ttgaaaaaca ggtcttagat 1500
ttgtttccat cagcccaggg ctgggaagtt acatgggtcca gtgtggtaaa gatcggacaa 1560
tccttgtacc gtgaggctcc tggaaatgat ccatttagac ctgaccagaa gacaccagtt 1620
aaaaatttct tcttgtctgg ctcttacacg aaacaggact acattgacag catggaagga 1680
gcaactcttt ctggcagggc aacagcagcc tacatctgtg gtgctggaga ggagttatta 1740
gccattcgaa agaagctcat tgtcgatcac agcgagaagg cctccgggat gggttcaaatg 1800
ttgcaaacaa gttaaaccctt gcaatcacca aagccagaat gccaaactat aggtttcagg 1860
gttataggcg attgtcaaca gtgatttcta taaatcttgt ttgcgcatta gatttggtgc 1920
tgcccttttt ttagccgtga cacatttgca catcgcatgt tcctatctta caattggcaa 1980
ccagggtgtct aagtggcata taggaatata tcaggcagcg tgtatgttca gttaaaccctt 2040
agcccgtcat tcgaatgatt gatgtcatat tgca 2074

```

<210> 8
 <211> 568
 <212> PRT
 <213> Triticum aestivum

<400> 8
 Met Ala Ala Thr Ser Cys Ala Leu Val Ser Ala Leu Val Val Gly Arg
 1 5 10 15
 Arg Arg Gly Pro Ser Cys Gln Arg Ala Ala Ala Ala Gly Val Val Arg
 20 25 30
 Cys Ser Leu Asp Ser Lys Val Ser Asp Met Ala Ile Asn Ala Pro Lys
 35 40 45
 Gly Leu Phe Pro Pro Glu Pro Glu His Tyr Arg Gly Pro Lys Leu Lys
 50 55 60
 Val Ala Ile Ile Gly Ala Gly Leu Ala Gly Met Ser Thr Ala Val Glu
 65 70 75 80
 Leu Leu Asp Gln Gly His Glu Val Asp Leu Tyr Asp Ser Arg Thr Phe
 85 90 95
 Ile Gly Gly Lys Val Gly Ser Phe Val Asp Lys His Gly Asn His Ile
 100 105 110
 Glu Met Gly Leu His Val Phe Phe Gly Cys Tyr Ser Asn Leu Phe Arg
 115 120 125
 Leu Met Lys Lys Val Gly Ala Asp Asn Asn Leu Leu Val Lys Glu His
 130 135 140
 Thr His Thr Phe Val Asn Lys Gly Gly Ile Val Gly Glu Leu Asp Phe
 145 150 155 160
 Arg Phe Pro Val Gly Ala Pro Leu His Gly Ile Gln Ala Phe Leu Arg
 165 170 175
 Thr Asn Gln Leu Lys Val Tyr Asp Lys Ala Arg Asn Ala Val Ala Leu
 180 185 190
 Ala Leu Ser Pro Val Val Arg Ala Leu Leu Asp Pro Asp Gly Ala Leu
 195 200 205

Gln Gln Val Arg Asp Leu Asp Asp Val Ser Phe Thr Asp Trp Phe Met
 210 215 220
 Ser Arg Gly Gly Thr Arg Glu Ser Ile Thr Arg Met Trp Asp Pro Val
 225 230 235 240
 Ala Tyr Ala Leu Gly Phe Ile Asp Cys Asp Asn Ile Ser Ala Arg Cys
 245 250 255
 Met Leu Thr Ile Phe Thr Leu Phe Ala Thr Lys Thr Glu Ala Ser Leu
 260 265 270
 Leu Arg Met Leu Lys Gly Ser Pro Asp Val Tyr Leu Ser Gly Pro Ile
 275 280 285
 Lys Lys Tyr Ile Thr Asp Arg Gly Gly Arg Phe His Leu Lys Trp Gly
 290 295 300
 Cys Arg Glu Val Leu Tyr Asp Lys Ser Pro Asp Gly Glu Thr Tyr Val
 305 310 315 320
 Lys Gly Phe Leu Ile Ser Lys Ala Thr Ser Ser Glu Ile Ile Lys Ala
 325 330 335
 Asp Ala Tyr Val Ala Ala Cys Asp Val Pro Gly Ile Lys Arg Leu Leu
 340 345 350
 Pro Ser Glu Trp Arg Glu Trp Asp Met Phe Asp Asn Ile Tyr Lys Leu
 355 360 365
 Asp Gly Val Pro Val Val Thr Val Gln Leu Arg Tyr Asn Gly Trp Val
 370 375 380
 Thr Glu Val Gln Asp Leu Glu Lys Ser Arg Gln Leu Gln Lys Ala Val
 385 390 395 400
 Gly Leu Asp Asn Leu Leu Tyr Thr Pro Asp Ala Asp Phe Ser Cys Phe
 405 410 415
 Ser Asp Leu Ala Leu Ser Ser Pro Ala Asp Tyr Tyr Ile Glu Gly Gln
 420 425 430
 Gly Ser Leu Ile Gln Ala Val Leu Thr Pro Gly Asp Pro Tyr Met Pro
 435 440 445
 Leu Pro Asn Glu Glu Ile Ile Ser Lys Val Glu Lys Gln Val Leu Asp
 450 455 460
 Leu Phe Pro Ser Ala Arg Gly Leu Glu Val Thr Trp Ser Ser Val Val
 465 470 475 480
 Lys Ile Gly Gln Ser Leu Tyr Arg Glu Ala Pro Gly Asn Asp Pro Phe
 485 490 495
 Arg Pro Asp Gln Lys Thr Pro Val Lys Asn Phe Phe Leu Ser Gly Ser
 500 505 510
 Tyr Thr Lys Gln Asp Tyr Ile Asp Ser Met Glu Gly Ala Thr Leu Ser
 515 520 525

Gly Arg Arg Thr Ala Ala Tyr Ile Cys Gly Ala Gly Glu Glu Leu Leu
530 535 540

Ala Ile Arg Lys Lys Leu Ile Val Asp His Ser Glu Lys Ala Ser Gly
545 550 555 560

Met Val Gln Met Leu Gln Thr Ser
565

<210> 9
<211> 2430
<212> DNA
<213> Oryza sativa

<400> 9
gcacgaggtt taaacccct ttgtcccttt cccaccgccc caaaaaccca cccctccct 60
gactcctccc ccgcagctt ccgcgcgtcc cctccgctcc cacgtcgccg cccgctcgtc 120
gtcgccgcgc acgctcttgc gtgcttattt gtcaaatacag atctgaatat aatttttagga 180
gttgcttcag catggatact ggctgcctgt catctatgaa cataactgga accagccaag 240
caagatcttt tgcgggacaa ctccctactc atagggtgctt cgcaagtagc agcatccaag 300
cactgaaaag tagtcagcat gtgagctttg gagtgaatc tcttgtctta aggaataaag 360
gaaaaagatt ccgtcggagg ctccgtgctc tacagggtgt ttgccaggac tttccaagac 420
ctccactaga aaacacaata aacttttttg aagctggaca actatcttca tttttcagaa 480
acagtgaaca accactaaa ccattacagg tcgtgattgc tggagcagga ttagctgggt 540
tatcaacggc aaaatatctg gcagatgctg gtcataaacc catattgctt gaggcaaggg 600
atgttttggg tggaaagata gctgcttggg aggatgaaga tggagattgg tatgaaactg 660
ggcttcatat cttttttgga gcttatccca acatacagaa ctgtgttggc gagcttggta 720
ttaatgatcg gttgcaatgg aaggaacact ccatgatatt tgccatgcc aacaagccag 780
gagaattcag ccggtttgat tttcctgaaa cattgcctgc acccttaaat ggaatatggg 840
ccatactaag aaacaatgaa atgctaactt ggccagagaa ggtgaagttt gctcttgac 900
ttttgccagc aatggttggg ggccaagctt atgttgaagc tcaagatggt tttactgttt 960
ctgagtgag gaaaaagcag ggtgttcctg atcgagtga cgatgaggtt ttcattgcaa 1020
tgtcaaaagg acttaatttc ataaatcctg atgagttatc catgcagtgc attctgattg 1080
ctttaaaccc atttcttcag gagaagcatg gttctaagat ggcatcttg gatggtaatc 1140
ctcctgaaag gttatgcatg cctattgttg accatgttcg ctctttgggt ggtgaggttc 1200
ggctgaattc tcgtattcag aaaatagaac ttaatcctga tggaaacagt aaacactttg 1260
cacttactga tggaaactca ataaactggag atgcttatgt ttttgcaaca ccagttgata 1320
tcttgaaagt tcttgtaact caagagtggg aagaaatata tgatagaaaa ctgaagaaca 1440
agttgggtgg agttcctgtt ataaatgttc atatatggtt tgatagaaaa ctgaagaaca 1440
catatgacca ccttcttttc agcaggagtt cacttttaag tgtttatgcg gacatgtcag 1500
taacttgcaa ggaatactat gatcaaaaac gttcaatgct ggagttggtc tttgctcctg 1560
cagaggaatg ggttggacgg agtgacactg aaatcatcga agcaactatg caagagctag 1620
ccaagctatt tcctgatgaa attgctgctg atcagagtaa agcaaagatt ctgaagtatc 1680
atgttgatgaa gacaccaaga tctgtttaca agactatccc ggactgtgaa ccttgccgac 1740
ctctgcaaag atcaccgatt gaagggttct atctagctgg tgactacaca aagcagaaat 1800
atgttgcttc gatggagggt gcagttctat ctgggaagct ttgtgctcag tctgtagtgg 1860
aggattataa aatgctatct cgtaggagcc tgaagaagct gcagcttgaa gttcctgttg 1920
cctcctagtt gtagtcagga ctattcccaa tgggtgtgtg gtcacatccc cctagtcagt 1980
ttttttctat ttagtggttg cccaactctc caccaattta cacatgatgg aacttgaaag 2040
atgcctatct tggcttctat atatttctgt aaagtgtgatt tgtggtttaa acctcggtgc 2100
gaattcggca cgaggtttta acgtgagatc gtgacgggtt tccaaccga gctcgaggtt 2160
attcggccaa agaagagcga caagatcaag ataatagaac gcaatttccg tggatactct 2220
ggaaagctca taggtataga tggttccgac ggcattgtga agcttgatga cacatacgaa 2280
gtcaagatct tagatatggt gattttggcc aaactggcga gttgagatgg atgtatat 2340
tgtacgagtc catcgtatag ccaaatgttt gtaacagata ccctttgtag tcaataaaaac 2400
gttttgttat ttaaaaaaaa aaaaaaaaaa 2430

<210> 10
<211> 578
<212> PRT
<213> Oryza sativa

```

<400> 10
Met Asp Thr Gly Cys Leu Ser Ser Met Asn Ile Thr Gly Thr Ser Gln
 1          5          10          15

Ala Arg Ser Phe Ala Gly Gln Leu Pro Thr His Arg Cys Phe Ala Ser
 20          25          30

Ser Ser Ile Gln Ala Leu Lys Ser Ser Gln His Val Ser Phe Gly Val
 35          40          45

Lys Ser Leu Val Leu Arg Asn Lys Gly Lys Arg Phe Arg Arg Arg Leu
 50          55          60

Gly Ala Leu Gln Val Val Cys Gln Asp Phe Pro Arg Pro Pro Leu Glu
 65          70          75          80

Asn Thr Ile Asn Phe Leu Glu Ala Gly Gln Leu Ser Ser Phe Phe Arg
 85          90          95

Asn Ser Glu Gln Pro Thr Lys Pro Leu Gln Val Val Ile Ala Gly Ala
100          105          110

Gly Leu Ala Gly Leu Ser Thr Ala Lys Tyr Leu Ala Asp Ala Gly His
115          120          125

Lys Pro Ile Leu Leu Glu Ala Arg Asp Val Leu Gly Gly Lys Ile Ala
130          135          140

Ala Trp Lys Asp Glu Asp Gly Asp Trp Tyr Glu Thr Gly Leu His Ile
145          150          155          160

Phe Phe Gly Ala Tyr Pro Asn Ile Gln Asn Leu Phe Gly Glu Leu Gly
165          170          175

Ile Asn Asp Arg Leu Gln Trp Lys Glu His Ser Met Ile Phe Ala Met
180          185          190

Pro Asn Lys Pro Gly Glu Phe Ser Arg Phe Asp Phe Pro Glu Thr Leu
195          200          205

Pro Ala Pro Leu Asn Gly Ile Trp Ala Ile Leu Arg Asn Asn Glu Met
210          215          220

Leu Thr Trp Pro Glu Lys Val Lys Phe Ala Leu Gly Leu Leu Pro Ala
225          230          235          240

Met Val Gly Gly Gln Ala Tyr Val Glu Ala Gln Asp Gly Phe Thr Val
245          250          255

Ser Glu Trp Met Lys Lys Gln Gly Val Pro Asp Arg Val Asn Asp Glu
260          265          270

Val Phe Ile Ala Met Ser Lys Ala Leu Asn Phe Ile Asn Pro Asp Glu
275          280          285

Leu Ser Met Gln Cys Ile Leu Ile Ala Leu Asn Arg Phe Leu Gln Glu
290          295          300

Lys His Gly Ser Lys Met Ala Phe Leu Asp Gly Asn Pro Pro Glu Arg
305          310          315          320

```

Leu Cys Met Pro Ile Val Asp His Val Arg Ser Leu Gly Gly Glu Val
 325 330 335
 Arg Leu Asn Ser Arg Ile Gln Lys Ile Glu Leu Asn Pro Asp Gly Thr
 340 345 350
 Val Lys His Phe Ala Leu Thr Asp Gly Thr Gln Ile Thr Gly Asp Ala
 355 360 365
 Tyr Val Phe Ala Thr Pro Val Asp Ile Leu Lys Leu Leu Val Pro Gln
 370 375 380
 Glu Trp Lys Glu Ile Ser Tyr Phe Lys Lys Leu Glu Lys Leu Val Gly
 385 390 395 400
 Val Pro Val Ile Asn Val His Ile Trp Phe Asp Arg Lys Leu Lys Asn
 405 410 415
 Thr Tyr Asp His Leu Leu Phe Ser Arg Ser Ser Leu Leu Ser Val Tyr
 420 425 430
 Ala Asp Met Ser Val Thr Cys Lys Glu Tyr Tyr Asp Gln Lys Arg Ser
 435 440 445
 Met Leu Glu Leu Val Phe Ala Pro Ala Glu Glu Trp Val Gly Arg Ser
 450 455 460
 Asp Thr Glu Ile Ile Glu Ala Thr Met Gln Glu Leu Ala Lys Leu Phe
 465 470 475 480
 Pro Asp Glu Ile Ala Ala Asp Gln Ser Lys Ala Lys Ile Leu Lys Tyr
 485 490 495
 His Val Val Lys Thr Pro Arg Ser Val Tyr Lys Thr Ile Pro Asp Cys
 500 505 510
 Glu Pro Cys Arg Pro Leu Gln Arg Ser Pro Ile Glu Gly Phe Tyr Leu
 515 520 525
 Ala Gly Asp Tyr Thr Lys Gln Lys Tyr Leu Ala Ser Met Glu Gly Ala
 530 535 540
 Val Leu Ser Gly Lys Leu Cys Ala Gln Ser Val Val Glu Asp Tyr Lys
 545 550 555 560
 Met Leu Ser Arg Arg Ser Leu Lys Ser Leu Gln Ser Glu Val Pro Val
 565 570 575

Ala Ser

<210> 11
 <211> 1420
 <212> DNA
 <213> Triticum aestivum

<400> 11
 gcacgaggtt tgctgagctt ggtattagt atcgcttgca atggaaggaa cactccatga 60
 tatttgccat gccaaacaaa ccaggagaat acagccgttt tgattttcca gagactttgc 120
 cggcgccctt aaatggagtg tgggccatac tgaaaaacaa tgaaatgctt acttggccgg 180

```

agaaggtgaa gtttgcattt gggcttctac ctgcaatgct tgggtggccaa gcttacgttg 240
aagctcaaga tggcttaact gtttcagaat ggatggaaaa gcagggtggt cctgatcggg 300
tcaacgacga ggttttttatt gcaatgtcca aggcactcaa tttcataaac cctgacgagt 360
tatccatgca gtgcattctg attgctctaa accgatttct ccaggagaag catggctcga 420
aaatggcatt cttggatggg aatcctcctg aaaggctatg catgcctatt gttaaccaca 480
ttcagtcttt ggggtggtgag gtccggctga attctcgtat tcagaaaatt gaactgaacc 540
cggacggaac agtgaagcac tttgcactta ctgatgggac tcaaataact ggagatgcat 600
atgttttttg agcaccagtt gatattcttca agcttcttgt accacaagag tggagagaga 660
tctcttattt caaaaggctg gataagttgg tgggagttcc tgtcatcaat gttcatatat 720
ggtttgacag aaaactgaaa aacacgtatg accaccttct tttcagcagg agttcacttt 780
taagcgttta tgcagacatg tcttttagcgt gcaaggagta ctatgatcca aaccgttcaa 840
tgctggagct ggtctttgct ccagcagagg aatggatcgg gcggagtgc accgaaatca 900
tcgaagcaac tatgctagag ctagccaagt tgtttcctga tgaaatcgct gctgaccaga 960
gtaaagcaaa gattcttaaa taccatgttg tgaagacacc gaggtccgtt tacaagactg 1020
tcccgaactg cgaaccttgc cgaccctgc aacgatcacc gatcgaaggg ttctatctgg 1080
ccggcgatta caaaaagcag aaataacctg cttccatgga ggggtgcggtt ttgtcaggga 1140
agttttgtgc tcagtccata gtgcaggatt ctaagatgct gtcccgagc agccaggaga 1200
gcctgcaatc cgaagccccg gtcgcctcca agttgtagct agttagcgcg attcaaaatt 1260
tttttggcgt ttcctatatg tcattgtcac attgtttag agtccaccag tgaattgagc 1320
tgacatccat attggaacta aaagggaat ttgtaaaaca aagaagacct tttgcagaag 1380
ggcaaaagtg ataaaaggaa tcttagatat caaaaaaaaa 1420

```

<210> 12
 <211> 411
 <212> PRT
 <213> Triticum aestivum

<400> 12
 Thr Arg Phe Ala Glu Leu Gly Ile Ser Asp Arg Leu Gln Trp Lys Glu
 1 5 10 15
 His Ser Met Ile Phe Ala Met Pro Asn Lys Pro Gly Glu Tyr Ser Arg
 20 25 30
 Phe Asp Phe Pro Glu Thr Leu Pro Ala Pro Leu Asn Gly Val Trp Ala
 35 40 45
 Ile Leu Lys Asn Asn Glu Met Leu Thr Trp Pro Glu Lys Val Lys Phe
 50 55 60
 Ala Ile Gly Leu Leu Pro Ala Met Leu Gly Gly Gln Ala Tyr Val Glu
 65 70 75 80
 Ala Gln Asp Gly Leu Thr Val Ser Glu Trp Met Glu Lys Gln Gly Val
 85 90 95
 Pro Asp Arg Val Asn Asp Glu Val Phe Ile Ala Met Ser Lys Ala Leu
 100 105 110
 Asn Phe Ile Asn Pro Asp Glu Leu Ser Met Gln Cys Ile Leu Ile Ala
 115 120 125
 Leu Asn Arg Phe Leu Gln Glu Lys His Gly Ser Lys Met Ala Phe Leu
 130 135 140
 Asp Gly Asn Pro Pro Glu Arg Leu Cys Met Pro Ile Val Asn His Ile
 145 150 155 160
 Gln Ser Leu Gly Gly Glu Val Arg Leu Asn Ser Arg Ile Gln Lys Ile
 165 170 175

Glu Leu Asn Pro Asp Gly Thr Val Lys His Phe Ala Leu Thr Asp Gly
 180 185 190
 Thr Gln Ile Thr Gly Asp Ala Tyr Val Phe Ala Ala Pro Val Asp Ile
 195 200 205
 Phe Lys Leu Leu Val Pro Gln Glu Trp Arg Glu Ile Ser Tyr Phe Lys
 210 215 220
 Arg Leu Asp Lys Leu Val Gly Val Pro Val Ile Asn Val His Ile Trp
 225 230 235 240
 Phe Asp Arg Lys Leu Lys Asn Thr Tyr Asp His Leu Leu Phe Ser Arg
 245 250 255
 Ser Ser Leu Leu Ser Val Tyr Ala Asp Met Ser Leu Ala Cys Lys Glu
 260 265 270
 Tyr Tyr Asp Pro Asn Arg Ser Met Leu Glu Leu Val Phe Ala Pro Ala
 275 280 285
 Glu Glu Trp Ile Gly Arg Ser Asp Thr Glu Ile Ile Glu Ala Thr Met
 290 295 300
 Leu Glu Leu Ala Lys Leu Phe Pro Asp Glu Ile Ala Ala Asp Gln Ser
 305 310 315 320
 Lys Ala Lys Ile Leu Lys Tyr His Val Val Lys Thr Pro Arg Ser Val
 325 330 335
 Tyr Lys Thr Val Pro Asn Cys Glu Pro Cys Arg Pro Leu Gln Arg Ser
 340 345 350
 Pro Ile Glu Gly Phe Tyr Leu Ala Gly Asp Tyr Thr Lys Gln Lys Tyr
 355 360 365
 Leu Ala Ser Met Glu Gly Ala Val Leu Ser Gly Lys Phe Cys Ala Gln
 370 375 380
 Ser Ile Val Gln Asp Ser Lys Met Leu Ser Arg Arg Ser Gln Glu Ser
 385 390 395 400
 Leu Gln Ser Glu Ala Pro Val Ala Ser Lys Leu
 405 410

<210> 13
 <211> 588
 <212> PRT
 <213> Capsicum annuum

<400> 13
 Met Ala Thr Cys Ser Ala Tyr Leu Cys Cys Pro Ala Thr Ser Ala Ser
 1 5 10 15
 Leu Lys Lys Arg Val Phe Pro Asp Gly Ser Ala Gly Phe Leu Phe Phe
 20 25 30
 Gly Gly Arg Arg Leu Ser Asn Arg Leu Val Thr Pro Lys Ser Val Ile
 35 40 45

Arg Ala Asp Leu Asn Ser Met Val Ser Asp Met Ser Thr Asn Ala Pro
 50 55 60
 Lys Gly Leu Phe Pro Pro Glu Pro Glu His Tyr Arg Gly Pro Lys Leu
 65 70 75 80
 Lys Val Ala Ile Ile Gly Ala Gly Leu Ala Gly Met Ser Thr Ala Val
 85 90 95
 Glu Leu Leu Asp Gln Gly His Glu Val Asp Ile Tyr Glu Ser Arg Thr
 100 105 110
 Phe Ile Gly Gly Lys Val Gly Ser Phe Val Asp Lys Arg Gly Asn His
 115 120 125
 Ile Glu Met Gly Leu His Val Phe Phe Gly Cys Tyr Asn Asn Leu Phe
 130 135 140
 Arg Leu Met Lys Lys Val Gly Ala Glu Lys Asn Leu Leu Val Lys Glu
 145 150 155 160
 His Thr His Thr Phe Val Asn Lys Gly Gly Glu Ile Gly Glu Leu Asp
 165 170 175
 Phe Arg Phe Pro Val Gly Ala Pro Leu His Gly Ile Asn Ala Phe Leu
 180 185 190
 Ser Thr Asn Gln Leu Lys Thr Tyr Asp Lys Ala Arg Asn Ala Val Ala
 195 200 205
 Leu Ala Leu Ser Pro Val Val Arg Ala Leu Val Asp Pro Asp Gly Ala
 210 215 220
 Leu Gln Gln Ile Arg Asp Leu Asp Ser Val Ser Phe Ser Asp Trp Phe
 225 230 235 240
 Met Ser Lys Gly Gly Thr Arg Ala Ser Ile Gln Arg Met Trp Asp Pro
 245 250 255
 Val Ala Tyr Ala Leu Gly Phe Ile Asp Cys Asp Asn Ile Ser Ala Arg
 260 265 270
 Cys Met Leu Thr Ile Phe Ala Leu Phe Ala Thr Lys Thr Glu Ala Ser
 275 280 285
 Leu Leu Arg Met Leu Lys Gly Ser Pro Asp Val Tyr Leu Ser Gly Pro
 290 295 300
 Ile Lys Lys Tyr Ile Ile Asp Lys Gly Gly Arg Phe His Leu Arg Trp
 305 310 315 320
 Gly Cys Arg Glu Val Leu Tyr Glu Thr Ser Ser Asp Gly Ser Met Tyr
 325 330 335
 Val Ser Gly Leu Ala Met Ser Lys Ala Thr Gln Lys Lys Ile Val Lys
 340 345 350
 Ala Asp Ala Tyr Val Ala Ala Cys Val Val Pro Gly Ile Lys Arg Leu
 355 360 365

Val Pro Gln Lys Trp Arg Glu Leu Glu Phe Phe Gly Asn Ile Tyr Lys
 370 375 380
 Leu Ile Gly Val Pro Val Val Thr Val Gln Leu Arg Tyr Asn Gly Trp
 385 390 395 400
 Val Thr Glu Leu Gln Asp Leu Glu Arg Ser Arg Gln Ser Lys Arg Ala
 405 410 415
 Thr Gly Leu Asp Asn Leu Leu Tyr Thr Pro Asp Ala Asp Phe Ser Cys
 420 425 430
 Phe Ala Asp Leu Ala Leu Ala Ser Pro Glu Asp Tyr Tyr Ile Glu Gly
 435 440 445
 Gln Gly Ser Leu Leu Gln Cys Val Leu Thr Pro Gly Asp Pro Tyr Met
 450 455 460
 Pro Leu Pro Asn Glu Glu Ile Ile Arg Arg Val Ser Lys Gln Val Leu
 465 470 475 480
 Ala Leu Phe Pro Ser Ser Gln Gly Leu Glu Val Thr Trp Ser Ser Val
 485 490 495
 Val Lys Ile Gly Gln Ser Leu Tyr Arg Glu Gly Pro Gly Lys Asp Pro
 500 505 510
 Phe Arg Pro Asp Gln Lys Thr Pro Val Glu Asn Phe Phe Leu Ala Gly
 515 520 525
 Ser Tyr Thr Lys Gln Asp Tyr Ile Asp Ser Met Glu Gly Ala Thr Leu
 530 535 540
 Ser Gly Arg Gln Ala Ser Ala Tyr Ile Cys Asp Ala Gly Glu Gln Leu
 545 550 555 560
 Leu Ala Leu Arg Lys Lys Ile Ala Ala Ala Glu Leu Asn Glu Ile Ser
 565 570 575
 Lys Gly Val Ser Leu Ser Asp Glu Leu Ser Leu Val
 580 585

<210> 14
 <211> 570
 <212> PRT
 <213> Zea mays

<400> 14
 Met Ala Ser Val Ala Ala Thr Thr Thr Leu Ala Pro Ala Leu Ala Pro
 1 5 10 15
 Arg Arg Ala Arg Pro Gly Thr Gly Leu Val Pro Pro Arg Arg Ala Ser
 20 25 30
 Ala Val Ala Ala Arg Ser Thr Val Thr Ser Pro Thr Trp Arg Gln Arg
 35 40 45
 Ser Gln Arg Leu Phe Pro Pro Glu Pro Glu His Tyr Arg Gly Pro Lys
 50 55 60

Leu Lys Val Ala Ile Ile Gly Ala Gly Leu Ala Gly Met Ser Thr Ala
 65 70 75 80
 Val Glu Leu Leu Asp Gln Gly His Glu Val Asp Leu Tyr Glu Ser Arg
 85 90 95
 Pro Phe Ile Gly Gly Lys Val Gly Ser Phe Val Asp Arg Gln Gly Asn
 100 105 110
 His Ile Glu Met Gly Leu His Val Phe Phe Gly Cys Tyr Ser Asn Leu
 115 120 125
 Phe Arg Leu Met Lys Lys Val Gly Ala Asp Asn Asn Leu Leu Val Lys
 130 135 140
 Glu His Thr His Thr Phe Val Asn Lys Gly Gly Thr Ile Gly Glu Leu
 145 150 155 160
 Asp Phe Arg Phe Pro Val Gly Ala Pro Leu His Gly Ile Gln Ala Phe
 165 170 175
 Leu Arg Thr Asn Gln Leu Lys Val Tyr Asp Lys Ala Arg Asn Ala Val
 180 185 190
 Ala Leu Ala Leu Ser Pro Val Val Arg Ala Leu Val Asp Pro Asp Gly
 195 200 205
 Ala Leu Gln Gln Val Arg Asp Leu Asp Asp Ile Ser Phe Ser Asp Trp
 210 215 220
 Phe Met Ser Lys Gly Gly Thr Arg Glu Ser Ile Thr Arg Met Trp Asp
 225 230 235 240
 Pro Val Arg Tyr Ala Leu Gly Phe Ile Asp Cys Asp Asn Ile Ser Ala
 245 250 255
 Arg Cys Met Leu Thr Ile Phe Thr Leu Phe Ala Thr Lys Thr Glu Ala
 260 265 270
 Ser Leu Leu Arg Met Leu Lys Gly Ser Pro Asp Val Tyr Leu Ser Gly
 275 280 285
 Pro Ile Lys Lys Tyr Ile Thr Asp Arg Gly Gly Arg Phe His Leu Arg
 290 295 300
 Trp Gly Cys Arg Glu Val Leu Tyr Glu Lys Ser Pro Asp Gly Glu Thr
 305 310 315 320
 Tyr Val Lys Gly Leu Leu Leu Thr Lys Ala Thr Ser Arg Glu Ile Ile
 325 330 335
 Lys Ala Asp Ala Tyr Val Ala Ala Cys Asp Val Pro Gly Ile Lys Arg
 340 345 350
 Leu Leu Pro Ser Glu Trp Arg Glu Trp Glu Met Phe Asp Asn Ile Tyr
 355 360 365
 Lys Leu Asp Gly Val Pro Val Val Thr Val Gln Leu Arg Tyr Asn Gly
 370 375 380

Trp Val Thr Glu Leu Gln Asp Leu Glu Lys Ser Arg Gln Leu Gln Arg
 385 390 395 400
 Ala Val Gly Leu Asp Asn Leu Leu Tyr Thr Ala Asp Ala Asp Phe Ser
 405 410 415
 Cys Phe Ser Asp Leu Ala Leu Ser Ser Pro Ala Asp Tyr Tyr Ile Glu
 420 425 430
 Gly Gln Gly Ser Leu Ile Gln Ala Val Leu Thr Pro Gly Asp Pro Tyr
 435 440 445
 Met Pro Leu Pro Asn Glu Glu Ile Ile Ser Lys Val Gln Lys Gln Val
 450 455 460
 Val Glu Leu Phe Pro Ser Ser Arg Gly Leu Glu Val Thr Trp Ser Ser
 465 470 475 480
 Val Val Lys Ile Gly Gln Ser Leu Tyr Arg Glu Ala Pro Gly Asn Asp
 485 490 495
 Pro Phe Arg Pro Asp Gln Lys Thr Pro Val Lys Asn Phe Phe Leu Ser
 500 505 510
 Gly Ser Tyr Thr Lys Gln Asp Tyr Ile Asp Ser Met Glu Gly Ala Thr
 515 520 525
 Leu Ser Gly Arg Arg Thr Ser Ala Tyr Ile Cys Gly Ala Gly Glu Glu
 530 535 540
 Leu Leu Ala Leu Arg Lys Lys Leu Leu Ile Asp Asp Gly Glu Lys Ala
 545 550 555 560
 Leu Gly Asn Val Gln Val Leu Gln Ala Ser
 565 570

<210> 15
 <211> 571
 <212> PRT
 <213> Zea mays

<400> 15
 Met Asp Thr Gly Cys Leu Ser Ser Met Asn Ile Thr Gly Ala Ser Gln
 1 5 10 15
 Thr Arg Ser Phe Ala Gly Gln Leu Pro Pro Gln Arg Cys Phe Ala Ser
 20 25 30
 Ser His Tyr Thr Ser Phe Ala Val Lys Lys Leu Val Ser Arg Asn Lys
 35 40 45
 Gly Arg Arg Ser His Arg Arg His Pro Ala Leu Gln Val Val Cys Lys
 50 55 60
 Asp Phe Pro Arg Pro Pro Leu Glu Ser Thr Ile Asn Tyr Leu Glu Ala
 65 70 75 80
 Gly Gln Leu Ser Ser Phe Phe Arg Asn Ser Glu Arg Pro Ser Lys Pro
 85 90 95

Leu Gln Val Val Val Ala Gly Ala Gly Leu Ala Gly Leu Ser Thr Ala
 100 105 110
 Lys Tyr Leu Ala Asp Ala Gly His Lys Pro Ile Leu Leu Glu Ala Arg
 115 120 125
 Asp Val Leu Gly Gly Lys Val Ala Ala Trp Lys Asp Glu Asp Gly Asp
 130 135 140
 Trp Tyr Glu Thr Gly Leu His Ile Phe Phe Gly Ala Tyr Pro Asn Ile
 145 150 155 160
 Gln Asn Leu Phe Gly Glu Leu Arg Ile Glu Asp Arg Leu Gln Trp Lys
 165 170 175
 Glu His Ser Met Ile Phe Ala Met Pro Asn Lys Pro Gly Glu Phe Ser
 180 185 190
 Arg Phe Asp Phe Pro Glu Thr Leu Pro Ala Pro Ile Asn Gly Ile Trp
 195 200 205
 Ala Ile Leu Arg Asn Asn Glu Met Leu Thr Trp Pro Glu Lys Val Lys
 210 215 220
 Phe Ala Ile Gly Leu Leu Pro Ala Met Val Gly Gly Gln Pro Tyr Val
 225 230 235 240
 Glu Ala Gln Asp Gly Leu Thr Val Ser Glu Trp Met Lys Lys Gln Gly
 245 250 255
 Val Pro Asp Arg Val Asn Asp Glu Val Phe Ile Ala Met Ser Lys Ala
 260 265 270
 Leu Asn Phe Ile Asn Pro Asp Glu Leu Ser Met Gln Cys Ile Leu Ile
 275 280 285
 Ala Leu Asn Arg Phe Leu Gln Glu Lys His Gly Ser Lys Met Ala Phe
 290 295 300
 Leu Asp Gly Asn Pro Pro Glu Arg Leu Cys Met Pro Ile Val Asp His
 305 310 315 320
 Ile Arg Ser Arg Gly Gly Glu Val Arg Leu Asn Ser Arg Ile Lys Lys
 325 330 335
 Ile Glu Leu Asn Pro Asp Gly Thr Val Lys His Phe Ala Leu Ser Asp
 340 345 350
 Gly Thr Gln Ile Thr Gly Asp Ala Tyr Val Cys Ala Thr Pro Val Asp
 355 360 365
 Ile Phe Lys Leu Leu Val Pro Gln Glu Trp Ser Glu Ile Thr Tyr Phe
 370 375 380
 Lys Lys Leu Glu Lys Leu Val Gly Val Pro Val Ile Asn Val His Ile
 385 390 395 400
 Trp Phe Asp Arg Lys Leu Asn Asn Thr Tyr Asp His Leu Leu Phe Ser
 405 410 415

Arg Ser Ser Leu Leu Ser Val Tyr Ala Asp Met Ser Val Thr Cys Lys
 420 425 430
 Glu Tyr Tyr Asp Pro Asn Arg Ser Met Leu Glu Leu Val Phe Ala Pro
 435 440 445
 Ala Asp Glu Trp Ile Gly Arg Ser Asp Thr Glu Ile Ile Asp Ala Thr
 450 455 460
 Met Glu Glu Leu Ala Lys Leu Phe Pro Asp Glu Ile Ala Ala Asp Gln
 465 470 475 480
 Ser Lys Ala Lys Ile Leu Lys Tyr His Ile Val Lys Thr Pro Arg Ser
 485 490 495
 Val Tyr Lys Thr Val Pro Asn Cys Glu Pro Cys Arg Pro Leu Gln Arg
 500 505 510
 Ser Pro Ile Glu Gly Phe Tyr Leu Ala Gly Asp Tyr Thr Lys Gln Lys
 515 520 525
 Tyr Leu Ala Ser Met Glu Gly Ala Val Leu Ser Gly Lys Leu Cys Ala
 530 535 540
 Gln Ser Ile Val Gln Asp Tyr Ser Arg Leu Ala Leu Arg Ser Gln Lys
 545 550 555 560
 Ser Leu Gln Ser Gly Glu Val Pro Val Pro Ser
 565 570



INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification ⁶ : C12N 15/82, 9/02, 5/10, G01N 33/50 // A01H 5/00	A3	(11) International Publication Number: WO 99/55888 (43) International Publication Date: 4 November 1999 (04.11.99)
---	-----------	---

(21) International Application Number: **PCT/US99/08746**
(22) International Filing Date: **21 April 1999 (21.04.99)**
(30) Priority Data:
60/083,042 **24 April 1998 (24.04.98)** **US**
(71) Applicant (for all designated States except US): **E.I. DU PONT DE NEMOURS AND COMPANY [US/US]; 1007 Market Street, Wilmington, DE 19898 (US).**
(72) Inventors; and
(75) Inventors/Applicants (for US only): **CAHOON, Rebecca, E. [US/US]; 2331 West 18th Street, Wilmington, DE 19806 (US). KINNEY, Anthony, J. [GB/US]; 609 Lore Avenue, Wilmington, DE 19809 (US). VOLLMER, Steven, J. [US/US]; 1019 Tweedbrook Road, Wilmington, DE 19810 (US). WENG, Zude [CN/US]; 909C Cloister Road, Wilmington, DE 19809 (US).**
(74) Agent: **MAJARIAN, William, R.; E.I. du Pont de Nemours and Company, Legal Patent Records Center, 1007 Market Street, Wilmington, DE 19898 (US).**

(81) Designated States: **AE, AL, AU, BA, BB, BG, BR, CA, CN, CU, CZ, EE, GD, GE, HR, HU, ID, IL, IN, IS, JP, KR, LC, LK, LR, LT, LV, MG, MK, MN, MX, NO, NZ, PL, RO, SG, SI, SK, SL, TR, TT, UA, US, UZ, VN, YU, ZA, ARIPO patent (GH, GM, KE, LS, MW, SD, SL, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).**

Published*With international search report.**Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.*

(88) Date of publication of the international search report:
13 April 2000 (13.04.00)

(54) Title: **CAROTENOID BIOSYNTHESIS ENZYMES**

(57) Abstract

This invention relates to an isolated nucleic acid fragment encoding a carotenoid biosynthetic enzyme. The invention also relates to the construction of a chimeric gene encoding all or a portion of the carotenoid biosynthetic enzyme, in sense or antisense orientation, wherein expression of the chimeric gene results in production of altered levels of the carotenoid biosynthetic enzyme in a transformed host cell.

```

**
SEQ ID NO:13  NASVAA--TTTAPALAPPRAPRPGTG-----LVPPR--RASAAVARSUTSPTR-
SEQ ID NO:14  NATCSATLCPATSBASLQGVFFDQAGLFTGGRLSHLV--TPKSVIRADINSHVSDM
SEQ ID NO:06  NAB---LTQCBATB-----LSAVFTTPTTKHSLKGLDAMVSDM
SEQ ID NO:08  NA--AT--SCALVSALVGRKR---G-----PBCQR--AAAAGVRCSLQSKVSDM
1 60

*****
SEQ ID NO:13  --GASNLFFPFEPIHNGKLRVALIGAGLGHSTAVELLDOGHEVLDYERPTFGKRG
SEQ ID NO:14  STNAPKGLTFPEPIHNGKLRVATIGAGLGHSTAVELLDOGHEVLDYERPTFGKRG
SEQ ID NO:06  SVNAPKGLTFPEPIHNGKLRVATIGAGLGHSTAVELLDOGHEVLDYERPTFGKRG
SEQ ID NO:08  ATNAPKGLTFPEPIHNGKLRVATIGAGLGHSTAVELLDOGHEVLDYERPTFGKRG
61 120

*****
SEQ ID NO:13  SFVDRQGNHENGILNVFFCCYSNLFALQKVGADNMLVKERTHTFVNKGOTIGELDPRF
SEQ ID NO:14  SFVDRQGNHENGILNVFFCCYSNLFALQKVGADNMLVKERTHTFVNKGOTIGELDPRF
SEQ ID NO:06  SFVDRQGNHENGILNVFFCCYSNLFALQKVGADNMLVKERTHTFVNKGOTIGELDPRF
SEQ ID NO:08  SFVDRQGNHENGILNVFFCCYSNLFALQKVGADNMLVKERTHTFVNKGOTIGELDPRF
121 180

*****
SEQ ID NO:13  PVGAPLNGIGATLTHQIAVYDKARNAVALALSPPVVALVDFGALQGVRLDDISFSDW
SEQ ID NO:14  PVGAPLNGIGATLTHQIAVYDKARNAVALALSPPVVALVDFGALQGVRLDDISFSDW
SEQ ID NO:06  PVGAPLNGIGATLTHQIAVYDKARNAVALALSPPVVALVDFGALQGVRLDDISFSDW
SEQ ID NO:08  PVGAPLNGIGATLTHQIAVYDKARNAVALALSPPVVALVDFGALQGVRLDDISFSDW
181 240

*****
SEQ ID NO:13  PMSGGYKRSITAMNDPVVALGFIQCDNISARQCHLITLITATKTEALLHLKGSPOV
SEQ ID NO:14  PMSGGYKRSITAMNDPVVALGFIQCDNISARQCHLITLITATKTEALLHLKGSPOV
SEQ ID NO:06  PMSGGYKRSITAMNDPVVALGFIQCDNISARQCHLITLITATKTEALLHLKGSPOV
SEQ ID NO:08  PMSGGYKRSITAMNDPVVALGFIQCDNISARQCHLITLITATKTEALLHLKGSPOV
241 300

*****
SEQ ID NO:13  YLSGPIKVIYDRCGRPHLRWCREVLYENSPDGETTVKGLLTKATSREI IKADATVAA
SEQ ID NO:14  YLSGPIKVIYDRCGRPHLRWCREVLYENSPDGETTVKGLLTKATSREI IKADATVAA
SEQ ID NO:06  YLSGPIKVIYDRCGRPHLRWCREVLYENSPDGETTVKGLLTKATSREI IKADATVAA
SEQ ID NO:08  YLSGPIKVIYDRCGRPHLRWCREVLYENSPDGETTVKGLLTKATSREI IKADATVAA
301 360

*****
SEQ ID NO:13  CVVPGIKELVPGKRELEFFGNITKLGIVPVVTVQLATKGVTVGLQLEKSRQKAVGL
SEQ ID NO:14  CVVPGIKELVPGKRELEFFGNITKLGIVPVVTVQLATKGVTVGLQLEKSRQKAVGL
SEQ ID NO:06  CVVPGIKELVPGKRELEFFGNITKLGIVPVVTVQLATKGVTVGLQLEKSRQKAVGL
SEQ ID NO:08  CVVPGIKELVPGKRELEFFGNITKLGIVPVVTVQLATKGVTVGLQLEKSRQKAVGL
361 420

*****
SEQ ID NO:13  DMLVTTADGFCFSDIALSFPADYIEGGGSLIGAVLTPGDPNPLPNEEII SKVKQV
SEQ ID NO:14  DMLVTTADGFCFSDIALSFPADYIEGGGSLIGAVLTPGDPNPLPNEEII SKVKQV
SEQ ID NO:06  DMLVTTADGFCFSDIALSFPADYIEGGGSLIGAVLTPGDPNPLPNEEII SKVKQV
SEQ ID NO:08  DMLVTTADGFCFSDIALSFPADYIEGGGSLIGAVLTPGDPNPLPNEEII SKVKQV
421 480

*****
SEQ ID NO:13  VELFPSEKGLVTVSSVVKIQGLYREAPGNDFFRPOKTPVWKTFLGSGTYTKDQYIDSM
SEQ ID NO:14  VELFPSEKGLVTVSSVVKIQGLYREAPGNDFFRPOKTPVWKTFLGSGTYTKDQYIDSM
SEQ ID NO:06  VELFPSEKGLVTVSSVVKIQGLYREAPGNDFFRPOKTPVWKTFLGSGTYTKDQYIDSM
SEQ ID NO:08  VELFPSEKGLVTVSSVVKIQGLYREAPGNDFFRPOKTPVWKTFLGSGTYTKDQYIDSM
481 540

*****
SEQ ID NO:13  EGATLSCGRASATICDAGZOLALNKKIAAEIHEISKVSLSDELIV
SEQ ID NO:14  EGATLSCGRASATICDAGZOLALNKKIAAEIHEISKVSLSDELIV
SEQ ID NO:06  EGATLSCGRASATICDAGZOLALNKKIAAEIHEISKVSLSDELIV
SEQ ID NO:08  EGATLSCGRASATICDAGZOLALNKKIAAEIHEISKVSLSDELIV
541 589
```

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece	ML	Mali	TR	Turkey
BG	Bulgaria	HU	Hungary	MN	Mongolia	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MR	Mauritania	UA	Ukraine
BR	Brazil	IL	Israel	MW	Malawi	UG	Uganda
BY	Belarus	IS	Iceland	MX	Mexico	US	United States of America
CA	Canada	IT	Italy	NE	Niger	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NL	Netherlands	VN	Viet Nam
CG	Congo	KE	Kenya	NO	Norway	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NZ	New Zealand	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	PL	Poland		
CM	Cameroon	KR	Republic of Korea	PT	Portugal		
CN	China	KZ	Kazakstan	RO	Romania		
CU	Cuba	LC	Saint Lucia	RU	Russian Federation		
CZ	Czech Republic	LI	Liechtenstein	SD	Sudan		
DE	Germany	LK	Sri Lanka	SE	Sweden		
DK	Denmark	LR	Liberia	SG	Singapore		
EE	Estonia						

INTERNATIONAL SEARCH REPORT

International Application No
PC1, JS 99/08746

A. CLASSIFICATION OF SUBJECT MATTER

IPC 6 C12N15/82 C12N9/02 C12N5/10 G01N33/50
//A01H5/00

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

IPC 6 C12N

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	SCOLNIK, P.A. AND BARTLEY, G.E.: "nucleotide sequence of zeta-carotene desaturase (Gen Bank U38550) from Arabidopsis thaliana" EMBL SEQUENCE DATA LIBRARY, 10 November 1995 (1995-11-10), XP002128769 heidelberg, germany accession no. U38550 ---	1-3,5
X	AL-BABILI, S., ET AL.: "a cDNA encoding for beta Carotene Desaturase (Accession no. AJ224683) from Narcissus pseudonarcissus" EMBL SEQUENCE DATA LIBRARY, 2 March 1998 (1998-03-02), XP002128770 heidelberg, germany accession no. AJ224683 --- -/--	1-3,5

☒ Further documents are listed in the continuation of box C.

☒ Patent family members are listed in annex.

* Special categories of cited documents :

- "A" document defining the general state of the art which is not considered to be of particular relevance
- "E" earlier document but published on or after the international filing date
- "L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
- "O" document referring to an oral disclosure, use, exhibition or other means
- "P" document published prior to the international filing date but later than the priority date claimed

- "T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
- "X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
- "Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.
- "&" document member of the same patent family

Date of the actual completion of the international search

26 January 2000

Date of mailing of the international search report

18.02.00

Name and mailing address of the ISA

European Patent Office, P.B. 5818 Patentlaan 2
NL - 2280 HV Rijswijk
Tel. (+31-70) 340-2040, Tx. 31 651 epo nl,
Fax: (+31-70) 340-3016

Authorized officer

Holtorf, S

INTERNATIONAL SEARCH REPORT

International Application No
PCT/US 99/08746

C (Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	ALBRECHT, M., ET AL.: "molecular cloning and functional expression in E. coli of a novel plant enzyme mediating zeta-carotene desaturation" FEBS LETTERS, vol. 372, 1995, pages 199-202, XP002112107 abstract, page 199; page 200, right column; Table 1; Fig. 1 ---	3
X	YAMAMOTO, K. AND SASAKI, T.: "rice cDNA from callus" EMBL SEQUENCE DATA LIBRARY, 6 August 1997 (1997-08-06), XP002128771 heidelberg, germany accession no. C27136 ---	1-3
X	US 5 539 093 A (FITZMAURICE WAYNE P ET AL) 23 July 1996 (1996-07-23) the whole document ---	6-13
X	LI, Z-H., ET AL.: "cloning and characterization of a maize cDNA encoding phytoene desaturase, an enzyme of the carotenoid biosynthetic pathway" PLANT MOLECULAR BIOLOGY, vol. 30, 1996, pages 269-279, XP002112111 the whole document ---	6-11
X	AL-BABILI, S., ET AL.: "a novel, soluble form of phytoene desaturase from Narcissus pseudonarcissus chromoplasts is Hsp70-complexed and competent for flavinylation, membrane association and enzymatic activation" THE PLANT JOURNAL, vol. 9, no. 5, 1996, pages 601-612, XP002128744 the whole document ---	6-10
X	& AL-BABILI, S., ET AL.: "a novel, soluble form of phytoene desaturase from Narcissus pseudonarcissus chromoplasts is Hsp70-complexed and competent for flavinylation, membrane association and enzymatic activation" EMBL SEQUENCE DATA LIBRARY, 18 April 1995 (1995-04-18), heidelberg, germany accession no. X78815.1 ---	6-10
	-/--	

INTERNATIONAL SEARCH REPORT

International Application No

PC1/US 99/08746

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	ALBRECHT, M., ET AL.: "biochemical characterization of purified zeta-carotene desaturase from Anabaena PCC7120 after expression in Escherichia coli" EUROPEAN JOURNAL OF BIOCHEMISTRY, vol. 236, 1996, pages 115-120, XP002112340 the whole document ---	16
X	SANDMANN, G., ET AL.: "a new non-radioactive assay of phytoene desaturase to evaluate bleaching herbicides" ZEITSCHRIFT FÜR NATURFORSCHUNG, vol. 51, July 1996 (1996-07), pages 534-538, XP002112339 the whole document ---	16
A	BARTLEY, G.E. AND SCOLNIK, P.A.: "Plant carotenoids: pigments for photoprotection, visual attraction, and human health" THE PLANT CELL, vol. 7, July 1995 (1995-07), pages 1027-1038, XP002112110 page 1027, right column; page 1030, right column; page 1035, Table 1 ---	1-5, 11-16
A	BREITENBACH, J., ET AL.: "a higher-plant type zeta-carotene desaturase in the cyanobacterium Synechocystis PCC6803" PLANT MOLECULAR BIOLOGY, vol. 36, March 1998 (1998-03), pages 725-732, XP002112112 abstract, page 725, right column; page 727; Fig. 2 + 4 ---	1-5, 11-16
A	BARTLEY, G.E., ET AL.: "molecular cloning and expression in photosynthetic bacteria of a soybean cDNA coding for phytoene desaturase, an enzyme of the carotenoid biosynthesis pathway" PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF THE USA, vol. 88, August 1991 (1991-08), pages 6532-6536, XP002128745 the whole document ---	1-5, 11-16
A	WO 95 34668 A (BIOSOURCE TECH INC) 21 December 1995 (1995-12-21) the whole document ---	1-16
A	WO 98 06862 A (SHEWMAKER CHRISTINE K ; CALGENE INC (US)) 19 February 1998 (1998-02-19) the whole document ---	1-16
	-/--	

INTERNATIONAL SEARCH REPORT

International Application No
PCT/US 99/08746

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
P,X	SHOEMAKER, R., ET AL.: "public soybean EST project" EMBL SEQUENCE DATA LIBRARY, 17 March 1999 (1999-03-17), XP002128772 heidelberg, germany accession no. AI495658 ---	1-3,5
P,X	LUO, R. AND WURTZEL, ET.: "untitled" EMBL SEQUENCE DATA LIBRARY, 6 January 1999 (1999-01-06), XP002128773 heidelberg, germany cited in the application accession no. AF047490 ---	1-3,5
P,X	VIGNESWARAN, A. AND WURTZEL, E.T.: "isolation of rice phytoene desaturase cDNA" EMBL SEQUENCE DATA LIBRARY, 6 January 1999 (1999-01-06), XP002128746 heidelberg, germany accession no. AF049356 -----	6,7,10

INTERNATIONAL SEARCH REPORT

In international application No.
PCT/US 99/08746

Box I Observations where certain claims were found unsearchable (Continuation of Item 1 of first sheet)

This International Search Report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☐ Claims Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:

2. ☐ Claims Nos.:
because they relate to parts of the International Application that do not comply with the prescribed requirements to such an extent that no meaningful International Search can be carried out, specifically:

3. ☐ Claims Nos.:
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box II Observations where unity of invention is lacking (Continuation of Item 2 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

see additional sheet

1. ☒ As all required additional search fees were timely paid by the applicant, this International Search Report covers all searchable claims.
2. ☐ As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.
3. ☐ As only some of the required additional search fees were timely paid by the applicant, this International Search Report covers only those claims for which fees were paid, specifically claims Nos.:
4. ☐ No required additional search fees were timely paid by the applicant. Consequently, this International Search Report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

Remark on Protest

- ☐ The additional search fees were accompanied by the applicant's protest.
- ☒ No protest accompanied the payment of additional search fees.

INTERNATIONAL SEARCH REPORT

International Application No. PCT/US 99/08746

FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210

1. Claims: 1-5,11-16 partially

Isolation of partial rice-specific cDNAs encoding zeta-carotene desaturases; namely SEQIDs 1 and 3, furthermore the corresponding deduced amino acid sequences SEQIDs 2 and 4.

2. Claims: 1-5,11-16 partially

Isolation of a partial soybean-specific cDNA encoding zeta-carotene desaturase; namely SEQIDs 5, furthermore the corresponding deduced amino acid sequence SEQIDs 6.

3. Claims: 1-5,11-16 partially

Isolation of a full-length wheat-specific cDNA encoding zeta-carotene desaturase; namely SEQIDs 7, furthermore the corresponding deduced amino acid sequence SEQIDs 8.

4. Claims: 6-10,11-16 partially

Isolation of a full-length rice-specific cDNA encoding Phytoene desaturase; namely SEQIDs 9, furthermore the corresponding deduced amino acid sequence SEQIDs 10.

5. Claims: 6-10,11-16 partially

Isolation of a full-length wheat-specific cDNA encoding Phytoene desaturase; namely SEQIDs 11, furthermore the corresponding deduced amino acid sequence SEQIDs 12.

INTERNATIONAL SEARCH REPORT

Information on patent family members

International Application No

PCT/JS 99/08746

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
US 5539093	A	23-07-1996	NONE
WO 9534668	A	21-12-1995	US 5922602 A 13-07-1999
		AU 710588 B 23-09-1999	
		AU 2653495 A 05-01-1996	
		CA 2193094 A 21-12-1995	
		EP 0804600 A 05-11-1997	
		JP 10501968 T 24-02-1998	
		ZA 9504451 A 05-02-1996	
WO 9806862	A	19-02-1998	AU 4058497 A 06-03-1998
			CN 1227609 A 01-09-1999
			EP 0925366 A 30-06-1999

